Investigation of the Zinc Binding Region of Prothymosin-alpha: A Spectroscopic and Thermodynamic Approach to Study Metal Binding in Intrinsically Disordered Proteins

Abstract

The goal of this study is to provide a deeper insight into how metals bind intrinsically disordered proteins (IDPs) by taking zinc binding to prothymosin-alpha as a case study. The involvement of metals in several diseases caused by (IDPs) has been noted long ago (Cu binding of α-synuclein in Parkinsons disease, prion aggregation due to Cu in madcow disease. etc.). However, there is still a lack of understanding of how metal binds IDPs. A deeper insight into metal binding by IDPs is essential for a full understanding of a proteome and to pave a way for drug development.

Prothymosin-alpha (Protα) is an acidic, natively unfolded, and highly conserved protein located mostly in the nucleus of eukaryotic cells. It consists of 110 amino acids out of which 53 residues are aspartic (D) and glutamic (E) acids thus it is highly acidic. The exact biological role of the protein is unknown. However, it has been shown to be involved in cell proliferation, chromatin remodeling, antiapoptotic activity etc. One of the most interesting characteristics of Protα is its ability to bind metal cations like Ca^{2+}, Mg^{2+}, Mn^{2+}, Zn^{2+}, Al^{2+}. In the background of all these cations, Protα selectively interacts with Zn^{2+} and undergoes a structural change plus the interactions of the protein with its binding partners are enhanced in the presence of Zn^{2+}. These features underline the importance of Zn^{2+} binding in Protα.

Very little information is available about Zn^{2+} binding to Protα. A recent NMR study has shown that the 48-110 segment of Protα is the Zn^{2+} binding region of the protein. Although the 48-110

region is shown to be the Zn^{2+} binding region, the central 50-89 segment, where most of the Zn^{2+} binding residues are located in the protein, was used for all the studies here.

To aid in the synthesis, purification, and characterization of highly negatively charged sequences like Protα(50-89)N50W new amino acid derivatives, 4-pyridylmethyl protected glutamic and aspartic acids were developed and their application was tested. Subsequently a circular dichroism (CD) spectroscopic study on Protα(50-89)N50W revealed that the peptide, like the full length protein, undergoes a structural change in the presence of Zn^{2+} and both the full protein and the peptide were saturated at 75 eq of Zn^{2+} indicating that the majority of Zn added binds in this region. The above study had also shown that selective binding of Zn^{2+} by Protα(50-89)N50W is due to sequence specificity and not solely because of electrostatic interactions between Protα(50-89)N50W and Zn^{2+}. Further the CD and differential scanning calorimetry (DSC) studies have shown that Protα(50-89)N50W undergoes a structural change similar to an alpha helix when heated from 20-80 ^{o}C.

To set the stage for determining the specific glutamic and aspartic residues in Protα that act as ligands for Zn^{2+}, a proof of principle for spin inversion recovery experiments has been provided using Zn^{2+} binding to EDTA as a model. This idea can further be extended to Protα(50-89)N50W to understand if Zn^{2+} binds to a specific set of residues or if several degenerate binding sites exist.

Investigation of the Zinc Binding Region of Prothymosin-alpha: A Spectroscopic and Thermodynamic Approach to Study Metal Binding in Intrinsically Disordered Proteins

By
SriRamya Garapati

A Dissertation

Submitted to the Faculty of

East Carolina University

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Chemistry

East Carolina University, North Carolina
2014

Investigation of the Zinc Binding Region of Prothymosin-alpha: A Spectroscopic and Thermodynamic Approach to Study Metal Binding in Intrinsically Disordered Proteins

By
SriRamya Garapati

APPROVED BY:

DIRECTOR OF DISSERTATION:

Colin Burns, Ph.D.

COMMITTEE MEMBER:

William Allen, Ph.D.

COMMITTEE MEMBER:

Art Rodriguez, Ph.D.

COMMITTEE MEMBER:

Anthony Kennedy, Ph.D.

COMMITTEE MEMBER:

Cindy Putnam Evans, Ph.D.

DIRECTOR OF INTERDISCIPLINARY PROGRAM IN BIOLOGICAL SCIENCES:

Terry West, Ph.D.

DEAN OF THE GRADUATE SCHOOL:

Paul Gemperline, Ph.D.

DEDICATED

To

My beloved parents and husband

Acknowledgements

I heart fully thank my advisor, Dr. Colin Burns, without whom I could not have made through graduate school. His encouragement, guidance and support helped me in every way to develop an understanding of the subject. His patience and simplicity has always surprised me. I am very grateful for all his help in this journey

I would like to extend my gratitude to Dr. Art Rodriguez and Dr. Anthony Kennedy who added their valuable suggestions and time to my research.

I also thank my fellow graduate students for all the support they have offered throughout this journey. A special thanks to my committee members and Department of Chemistry for their assistance and time.

Lastly, I offer my regards to all of those who supported me in any respect during the completion of the project.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF SEQUENCES

LIST OF SCHEMES

LIST OF ABBREVIATIONS AND SYMBOLS

IDP: Intrinsically Disordered protein

Protα: Prothymosin-alpha

Tα-1: Thymosin-alpha1

CD: Circular dichroism

SAXS: Small angle X-ray scattering

CREB: cAMP response element–binding protein

FRET: Fluorescence resonance energy transfer

mM: millimolar

μM: micromolar

Glu: glutamic acid

Asp: Aspartic acid

Fmoc: 9-Fluorenylmethoxycarbonyl

A: Absorbance

ε: Extinction coefficient

b: Path length

c: Concentration

SPPS: Solid phase peptide synthesis

HBTU: O-(benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HOBt: 1-hydroxybenzotriazole

HATU: O-(7-azabenzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate

TIS: Triisopropylsilane

DMF: *N*,*N*-Dimethylformamide

DIEA: *N*,*N*-Diisopropylethylamine

TFA: Trifluoroacetic acid

DCM: Dichloromethane

MS: Mass spectrometry

HPLC: High performance liquid chromatography

MWCO: Molecular weight cut

MΩ cm-1: megaohm inverse centimeters

L: Liter

ESI-MS: Electrospray Ionization-Mass spectrometry

mg: milligram

SAX: Strong anion exchange

S/N: Signal to noise

m/z: mass to charge ratio

OBzl: Benzyl ester

Da: Daltons

tBu: tertiary butyl

NOE: Nuclear overhauser effect

PolyE: polyglutamic acid

OMePyr: 4-pyridylmethyl

DSC: Differential scanning calorimetry

Chapter 1 Introduction

Section 1.1 Natively Unfolded Proteins

The term "natively unfolded" protein is used to refer proteins at neutral pH that have been experimentally shown or predicted to lack ordered structure *(1)*.

Natively unfolded proteins are classified into two categories, those that are completely disordered throughout their length (often called as natively unfolded proteins, NUP) and those that contain a disordered region of at least 30-50 consecutive residues. The latter are called intrinsically disordered proteins (IDP) *(1, 2)*.

Various terms have been used to describe NUP's or IDP's, including: flexible *(3)* mobile *(4)*, partially folded, natively denatured *(5)*, natively unfolded *(6)* intrinsically unstructured *(7)*, rheomorphic *(8)*, natively disordered *(9)*, and intrinsically disordered *(2*, *18)*. None of these terms or combinations is completely appropriate. Flexible, mobile, and partially folded have the longest histories and most extensive use; however, all three of these eight terms have been borrowed from protein folding or crystallography and are not directly related to the biologically active proteins that normally exist as structural ensembles. For example, partially folded is often used to describe transient intermediates involved in protein folding. With regard to the term natively, it is difficult to know whether a protein is in its native state. Even when under apparently physiological conditions, a protein might fail to acquire a specific 3-D structure due to the absence of a critical ligand or because the crowded conditions inside the cell are needed to promote folding. Because of such uncertainties, "intrinsically" is often chosen over "natively" *(9)*. Unfolded and denatured are often used interchangeably, so the oxymoron "natively denatured" has a certain appeal but has not gained significant usage. Unfolded and unstructured

both imply lack of backbone organization, but natively disordered proteins often have regions of secondary structure, sometimes transient and sometimes persistent.

The term “intrinsically disordered proteins” (IDP) will be used throughout the document to refer to proteins that contain a disordered region of at least 30-50 consecutive residues.

Prediction of disorder in proteins

Using predictor programs it was determined that more than one-third of the proteins are disordered *(10)*. Amino acid analysis of approximately 100 natively unfolded proteins has led to the prediction that disorder is related to protein sequence *(11)*. In an attempt to understand the relationship between sequence and disorder, Dunker and coauthors have developed several neuronal network predictors. They assumed that if a protein structure has evolved to have a functional disordered state then a propensity for disorder might be predictable from its amino acid sequence and composition. By analyzing many proteins (~100) that lack secondary structure it was found that the common features include lack of sequence complexity, have high net charge/low hydrophobicity, and large hydrodynamic volume. Knowing about these three features gives predictive power.

1. Natively unfolded proteins are analyzed for sequence complexity. Sequences of natively unfolded proteins/regions may be essentially degenerate (contain same amino acid/set of amino acids repeatedly) and be less complex *(12)*.
2. The presence of high net charge and low hydrophobicity is another feature of disorder. It is an important prerequisite for the absence of compact structure in proteins under physiological conditions. Figure 1.1 shows comparison of mean hydrophobicity and mean net charge for a set of 275 folded and 102 intrinsically disordered proteins. The line

stands as a barrier between hydrophobicity of folded and unfolded proteins. Disordered proteins showed less hydrophobicity than the folded ones.

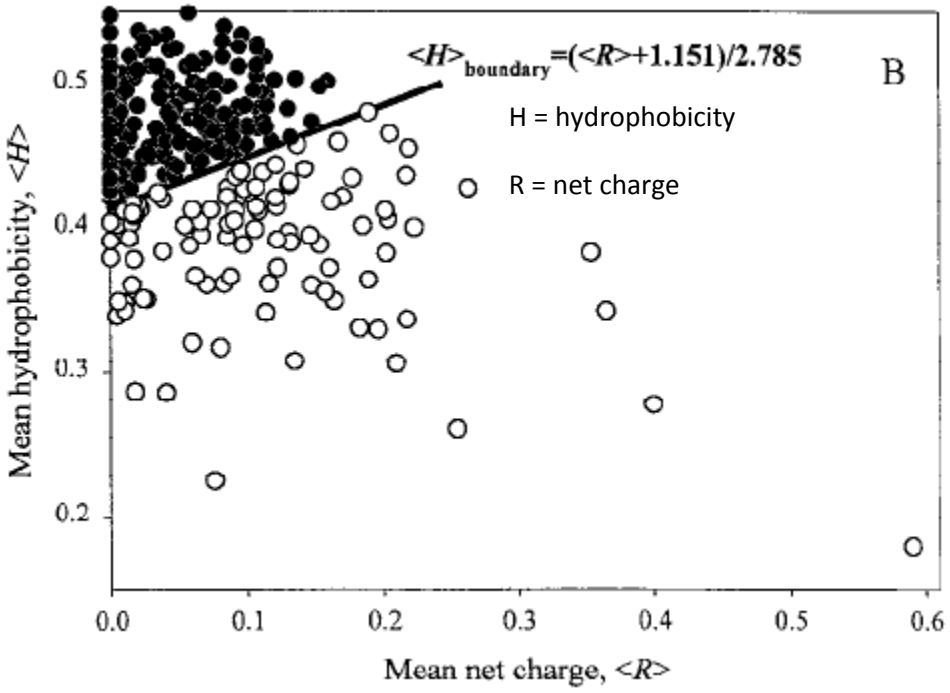


Figure 1.1: Comparison of the mean hydrophobicity and the mean net charge for a set of 275 folded (black circles) and 102 natively unfolded proteins (open circles) *(12)*.

3. Hydrodynamic volume for natively unfolded proteins was calculated from the Stokes equation.

$$D = \frac{KT}{6\pi\acute{\eta}R_H}$$

Measuring the diffusion coefficient (D) of proteins in solution allows one to calculate the hydrodynamic radius by using stokes equation. IDP's are found to have higher hydrodynamic volume compared to folded proteins.

Apart from the conformational analysis studies several predictor programs have evolved to study protein disorder *(13)*. Sequence disorder of a protein can be predicted using computational

methods. The first predictor program of ID regions was called as Neuronal Network Predictor. It was reported in 1997 *(14)*. The program works on local amino acid composition, hydropathy, and flexibility bases. The predictive model was later extended to the VLXT predictor *(15)* by a descriptive prefix, Predictor of Natural Disordered Regions (PONDR), giving PONDR VLXT. In 2000, it was noticed that natively unfolded proteins can be separated from ordered proteins by considering their average net charge and hydropathy *(16)*. This observation led to the development of a simple binary classifier, the charge hydropathy plot (CH-plot) *(16)*. The CH-plot works on the analysis of global sequence composition and classifies the entire protein as compact or natively unfolded. Another binary classifier is the cumulative distribution functions (CDF) analysis of disorder scores, which separates ordered and disordered sequences based on the per-residue disorder score retrieved by PONDR VLXT, and the optimal boundary *(17*, *18)*. This method summarizes the per-residue predictions by plotting PONDR scores against their cumulative frequency, which allows ordered and disordered proteins to be distinguished based on the distribution of prediction scores. Later, more sophisticated methods based on various statistical and machine learning techniques were published *(19, 20)*.

Biophysical characterization of IDPs

The demand for structural and molecular dynamics data of IDPs is growing. A need for deeper insight into the dynamics of IDP's necessitated the development of new experimental techniques *(21)*. Additionally, computational techniques have been invented to interpret experimental data to obtain accurate structural models of IDPs. Algorithms such as flexible-meccano *(22)* have been created specifically for the generation of ensembles of conformations, the form of representation usually chosen for the molecular description of IDPs *(23, 24)*.

Although several experimental techniques and computational analyses that provide insight into the dynamics of IDP's have been developed to obtain the structural models of proteins, NMR stands as the first technique to probe protein structure propensities and dynamics. However, in the case of IDP's the lack of dispersion of proton resonances and severe signal-overlap causes problems in NMR of IDPs, while intramolecular motions cause slower relaxation rates and allow the acquisition of spectra with narrow lines *(25, 26, 27)*. Advances in methodology, sample preparation techniques, labeling schemes, and NMR technology helped to overcome the problem of signal-overlap, thus enabling the study of more, larger IDPs by NMR spectroscopy *(26)*. However, intrinsically disordered regions are often part of large multi-domain proteins, connecting well-structured enzymatic domains and should be investigated in context with the super tertiary structure of the protein *(27)*. In two recent studies, the super tertiary structure of the membrane associated guanylate kinase (MAGuK) PSD-95, a large scaffold protein with a number of disordered linkers connecting protein-binding domains, was resolved by employing single molecule FRET (fluorescence resonance energy transfer) *(28)*, and a combination of NMR and small angle X-ray scattering (SAXS) *(29)*. This underlines the importance of complementary approaches with other biophysical techniques to obtain a complete picture of the complicated conformational ensembles.

Experimental determination of lack of appreciable secondary structure requires combination of different biophysical methods such as CD, NMR, dynamic light scattering (to get the hydrodynamic radius), and SAXS. If the protein exhibits certain spectral features (random coil CD, large hydrodynamic volume, lacks a maximum in the Kratky plot) it can be classified as an IDP.

1. Secondary structure: Unfolded polypeptide chains are characterized by very specific shapes of their far-UV CD spectrum, with an intensive minimum in the vicinity of 200 nm and an ellipticity close to zero in the vicinity of 222 nm. A random coil-like shape of far-UV CD spectrum has been reported for ~100 proteins, which is almost threefold larger than the number of proteins shown to be unfolded in accordance with their hydrodynamic dimensions *(11)*. The far-UV CD spectra of α-synuclein, prothymosin-α, phosphodiesterase γ-subunit and the caldesmon 636-771 fragment are shown in Figure 1.2 as typical representatives of the family of natively unfolded proteins. One can see that these proteins (as well as all other intrinsically unstructured proteins, whose far-UV CD spectra were studied) possess distinctive far-UV CD spectra with characteristic deep minima in the vicinity of 200 nm, and relatively low molar ellipticity at 220 nm. The analysis of these spectra indicates low content of ordered secondary structure (α-helices and β-sheets). This is also confirmed by the Fourier-transform infrared (FTIR) analysis of secondary structure composition of natively unfolded proteins, such as tau protein *(5)*, α-synuclein *(6, 30)*, casein *(31)*, and cAMP-dependent protein kinase inhibitor *(32)*. Importantly, even the caldesmon 636-771 fragment, which was shown to have hydrodynamic properties typical of the pre-molten globule possesses far-UV CD characteristic of essentially disordered polypeptide chain.

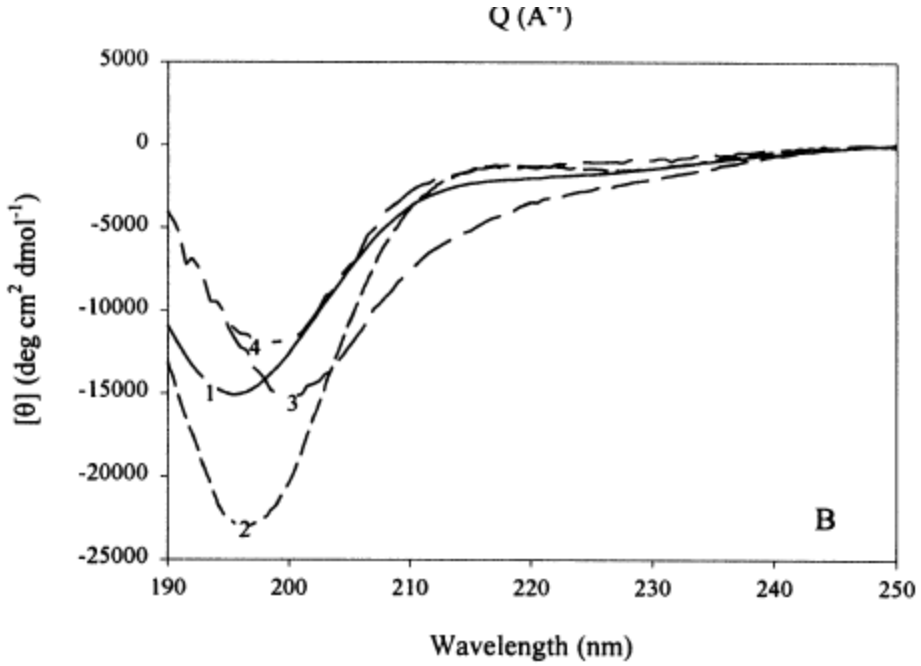


Figure 1.2: Far-UV CD spectra of intrinsically unordered proteins, α-synuclein (1), prothymosin-α (2), caldesmon 636-771 fragment (3) and phosphodiesterase c-subunit (4) *(11)*.

2. Another very important structural parameter to classify a protein as folded or unfolded protein is the degree of globularization that reflects the presence or absence of tightly packed core in the protein molecule. This information could be extracted from the analysis of small angle X-ray scattering (SAXS) data (Kratky plot), whose shape is sensitive to the conformational state of the scattering protein molecules *(33, 34)*. It has been shown that a scattering curve in the Kratky plot has a characteristic maximum when the globular protein is in the native state or in the molten globule state (i.e. has a globular structure). If a protein is completely unfolded or in a pre-molten globule conformation, such a maximum will be absent on the respective scattering curve *(33, 35, 36)*. Figure 1.3 compares the Kratky plots of three natively unfolded proteins (α-synuclein, prothymosin-α and the caldesmon 636-771 fragment) with that of the rigid globular protein SNase. One can see that intrinsically disordered proteins give Kratky plots without maxima typical of folded conformations of

globular proteins. Thus, these three natively unfolded proteins are characterized by the absence of globular structure, or, in other words, they do not have a tightly packed core under physiological conditions in vitro.

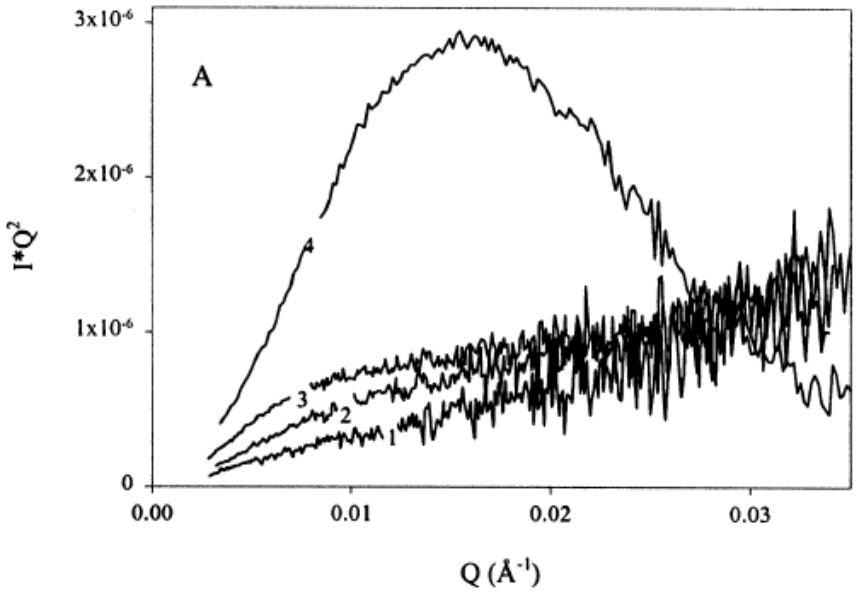


Figure 1.3: Kratky plots of SAXS data for natively unfolded a-synuclein (1), prothymosin-α (2) and caldesmon 636-771 fragment (3). The Kratky plot of native globular SNase is shown for comparison *(4, 11)*.

Structure function paradigm

It is widely believed that the function of a protein is related to its 3D structure. Though the presence of unfolded proteins or unfolded segments in proteins has been known, they were not a focus of much research until a key publication by Wright and Dyson came out in 1999 *(7)*. Wright and Dyson have made a comprehensive survey of experimental works on proteins, including NMR measurements revealing that various proteins interact with other proteins via intrinsically disordered (ID) regions, thus bringing into light the importance of natively unfolded proteins. Using conformational analysis and several prediction programs approximately 100

unfolded proteins have been analyzed experimentally *(11)*. In short, prediction of disorder has given an alternative view to the structure-function paradigm.

Here we explore the distinction between the standard and alternative view of structure-function paradigm. The standard structure-function paradigm arose from the study of enzymes. The original lock and key proposal *(9)* was based on distinctive substrate recognition by an enzyme. Enzymes bring substrates together in an optimal orientation to facilitate the making or breaking of chemical bonds through stabilization of the transition state. To achieve this, the enzyme must possess a unique active site where the substrates bind. The positioning of the key residues that form the active site is critical to the proper functioning of the enzyme. The active site is usually a 3D cleft in the protein that exists as a result of the overall 3D structure of the protein. Thus, in short, the evolution of ordered structure in proteins was likely reinforced by or directly resulted from the importance of enzyme catalysis.

The functional importance of being disordered has been extensively analysed. IDPs are not well suited to enzymatic activity as they do not have the propensity to form structures which are required to form active sites. For most of the natively disordered proteins studied *(37, 38)*, the function of the disorder is for signalling, regulation. The majority of disordered proteins undergo disorder to order transition upon binding its partner such as a metal or another protein. The free energy required to bring about the disorder to order transition takes away from the interfacial, contact free energy, with the net result that a highly specific interaction can be combined with a low net free energy of association *(2)*. In addition, a disordered protein can have multiple partners and can associate with several targets.

Given the above information, a new protein structure-function paradigm was proposed *(9)*. Simply put, a two-pathway paradigm, with sequence to 3-D structure to function for catalysis

(Figure 1.4A) and sequence to native disorder to function for signalling and regulation (Figure 1.4B).

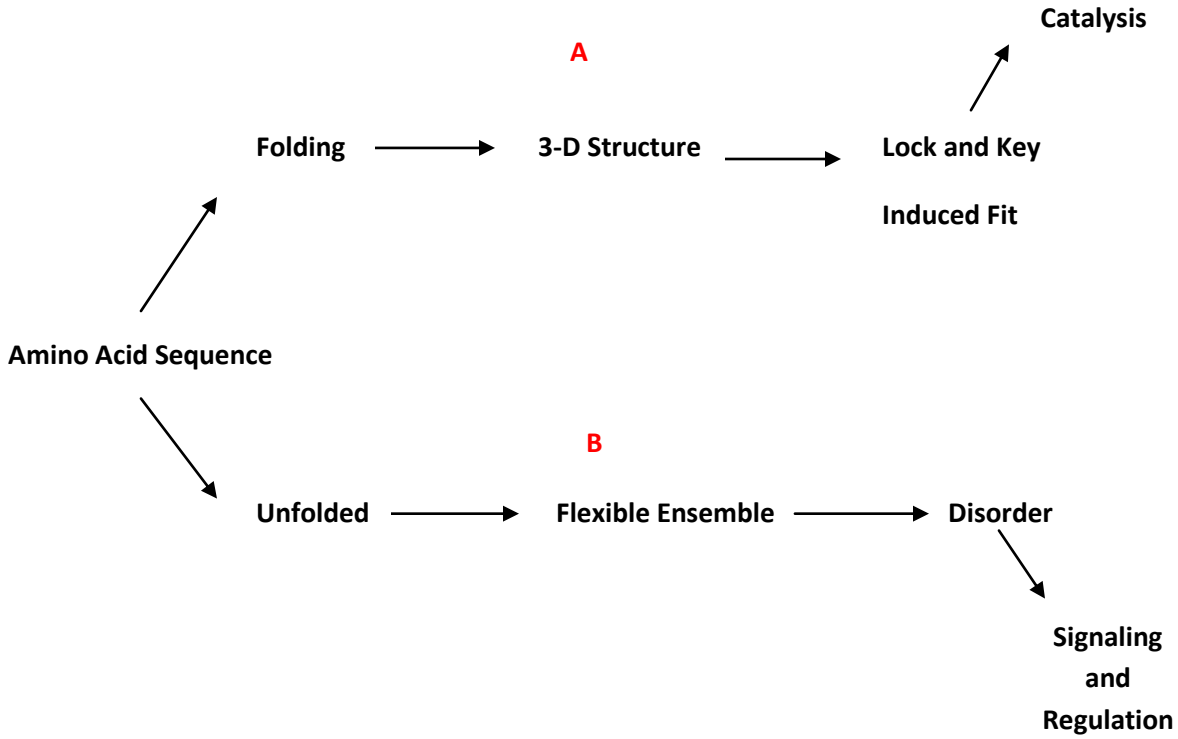


Figure 1.4: A) Protein structure-function paradigm showing A) folded protein to function (catalysis) B) Alternative view (IDP to signaling and regulation)

The next important topic of interest is metal binding of intrinsically disordered proteins. Correlating the facts that one-third of the proteins are disordered and more than one-third of the all the proteins require metal co-factors to carry out their function *(39, 40)* raises in the question about the extent of overlap of IDPs with metalloproteins. All these sum up as key reasons to

study the importance of metals in the functioning of natively unfolded proteins by biophysically characterizing metallo-IDPs.

Metal ions in folded proteins versus intrinsically disordered proteins

In folded proteins metals are known to act as cofactors. For example, heme proteins, which contain a porphyrin ring, coordinate to iron and this coordination is necessary for the transport of oxygen. The binding of metal ions by proteins can stabilize the structure, as is Zinc finger proteins, for example metal ion-binding proteins increase the structural stability and serve as a cross-linking agent to bring reactive groups into the correct orientation for reaction. They can serve as redox centers for catalysis (e.g., haemiron centers) or as electrophilic reactants in catalysis *(20)*. Metals can also play a regulatory role in proteins *(41)*.

In IDP's metals are mostly known to induce a structural change. For example, prothymosin-alpha is known to attain a compact conformation in the presence of Zn^{2+}. Zn^{2+} binding induces partial folding of proteins such as thymosin α1 *(42)*, Prothymosin-α *(43)* human sperm Protamines P2 and P3 (Gatewood et al. 1996), and phosphodiesterase γ-subunit *(11)*. Human α-synuclein which is also a natively unfolded protein was also shown to be partially folded in the presence of several divalent and trivalent metal ions. Unlike the metal bound folded proteins very little information is known about metallo-IDPs. For example, prothymosin-α is known to bind Zn^{2+} but the importance of Zn^{2+} binding in prothymosin-α's function is yet to be revealed. The protein α-synuclein is known to bind metal ions but the role of the protein interaction with the metal in Parkinsons disease is still under research *(44)*.

Why should we study IDP's and metal binding in IDP's?

Despite the lack of a stable secondary and tertiary structure, IDPs play important roles in critical, mostly regulatory, cellular functions such as differentiation, transcription regulation, DNA condensation, mRNA processing, and apoptosis *(27, 45)*. These functions are frequently modulated by post-translational modification such as phosphorylation or glycosylation. Numerous IDPs have been connected to diseases; examples include α-synuclein in Parkinson's, tau protein in Alzheimer's, or the role of tumor suppressor p53 in cancer formation *(46)*. The biological importance of IDPs becomes evident with the severity of the pathological conditions associated with malfunctions of IDPs. Due to their involvement in disease, IDPs became of interest for drug discovery, and strategies to create molecules that would inhibit disorder-based interactions of IDPs have been developed *(25)*. Depending on the pathogenesis, drugs have been designed to either inhibit binding, like in the case of the oncoprotein c-Myc *(47)*, or to block amyloid aggregation, the key event in neurodegenerative disease *(48)*. But regardless of the desired molecular effect of the drug, atomic-scale structural characterization of the target allows rational drug design and the discovery of more effective treatments *(25, 48)*. Therefore, understanding the molecular features of IDP's and biophysical characterization of metal binding to these proteins has gained attention.

1.2 General Introduction about Prothymosin-α (Protα)

Numerous examples of natively unfolded proteins and protein domains exist *(7, 30, 49)* and prothymosin alpha (Protα) is one among them *(50, 51)*. It is a highly conserved protein in eukaryotes. The highly conserved nature of the protein suggests the essential role for this protein at cellular level *(52)*. It is present in numerous tissues in mammals. It is synthesized in cytoplasm and migrates to the nucleus *(53)*.

It is a highly acidic protein due to the fact that 52% of the residues are aspartic and glutamic acids. Most of the acidic residues are clustered in the central region (Figure 1.5).

S D A A V D T S S E I T T K D L K E K K E V V E E A E N G R D A P A N G N A E N E E N G E Q E A D **N E V D E E E E E G G E E E E E E E E G D G E E E D G D E D E E A E S A T G K R** A A E D D E D D D V D T K K Q K T D E D D

Figure 1.5: Amino acid sequence of human prothymosin-alpha. The sequence is highly acidic with most of the acidic residues (glutamic and aspartic acids) located in the central segment of the sequence.

Protα is highly hydrophilic in nature with a low isoelectric point (pI = 3.55) *(54)*. The protein does not contain aromatic or sulfur containing amino acids and therefore does not absorb light at 280 nm. It has four valines and one leucine while the rest of the residues are polar or small like glycine or alanine.

Posttranslational modifications of Protα have been reported including unusual phoshorylation at glutamic acid residues and proteolytic cleavage of the N-terminus by lysosomal asparaginyl endopeptidase producing a 28 amino acid bioactive peptide called thymosin-α 1 (Tα1) *(55).*

Protα is classified as a Thymosin with nuclear localization sequence *(56).* Three stretches of basic residues are located in the sequence. The motif KEKKE located at the N-terminus apparently does not possess nuclear import activity *(57),* whereas the motifs KR and KKQK located in the C-terminus of the proteins are potent signals for nuclear localization *(58)*.

At pH 7.5, by using a combination different techniques discussed in section 1.1 Protα has been found to have no regular secondary structure, and its hydrodynamic dimensions are typical of a random coil *(43)*. Through the use of circular dichrosim (CD) it has been shown that Protα in solution adopts a random coil like conformation *(59).* The structural properties and conformational stability of recombinant human prothymosin-α were characterized at neutral and acidic pH by gel filtration, small angle X-ray scattering (SAXS), and ANS fluorescence. At neutral pH recombinant human Protα possesses all the properties of the unfolded state, whereas lowering the pH results in the transformation of the protein into a partially folded conformation *(57, 58, 59, 9).* The topic will be elaborated in section 1.4.

The protein is acetylated at the N-terminus to protect against proteolytic degradation. The central region of Protα carries most of the acidic residues and is thought to be a metal binding region capable of binding zinc *(60)*. Metal binding of this protein will be discussed in later sections.

1.3 Biological role of Protα

Since the discovery of Protα, its exact biological role has been obscure *(61)*. It is a small acidic protein that has been implicated in a number of diverse biological activities. At first Protα was isolated from rat thymus and considered as a thymic factor with a hormone-like role in the maturation of T-lymphocytes *(62)*. It has been reported to be involved in numerous activities including cell proliferation *(52)*, translocation, chromatin remodeling *(63)*, DNA packaging. It is also known to have an immunomodulatory role *(64)*.

It has been shown that expression of Protα is elevated in proliferating tissues *(63, 61, 65)* and over expression of Protα results in shortening of G1 phase of the cell cycle. These findings indicate that Protα must be a limiting factor necessary for the progression of cell cycle *(63, 73)*. Because of its role in cell cycle progression it has been used as a proliferation marker in various tumoral processes. Malignant tissues show higher levels of Protα expression compared to healthy tissues surrounding the area *(66, 67)*. Due to this finding it has been established as a marker for breast cancer *(68)*. Protα has been of prognostic value as it is associated with factors that describe the metastatic potential of the tumor and the risk of death *(67)*. This protein also serves as a marker for hepatocarcinoma.

The location of Protα in the cell nucleus, its acidic nature which resembles some chromatin/chromosomal proteins, and its structural similarity to nucleoplasmin, a chromatin remodeling protein in Xenopus, suggests that Protα is involved in chromatin organization *(53)*. Further, the interaction of Protα with histone H1 in vitro also supports this idea.

Protα participates in DNA packaging *(51, 69)*. It is involved in translocation of STAT3 which is a transcription factor that dimerizes and translocates into the nucleus to activate transcription of

IFN-stimulated genes. The nuclear targeting of Protα to translocate STAT3 is mediated by a bipartite nuclear localization signal (KR and KKQK) positioned close to the C-terminus *(70)*. (Figure 1.5)

Immunomodulatory Role:

As mentioned above, Protα has an immunomodulatory role. It has been recently discovered that exogenous Protα acts as anti-HIV agent. $CD8^+$ T lymphocyte cells secrete soluble molecules capable of inhibiting HIV-1 replication in vitro and Protα was identified in fractions of $CD8^+$ T cell secretions. The protein is known to suppress HIV replication by type 1 interferon (IFN) induction *(64, 71-73)*. In an effort to identify the anti-HIV domain(s) of Protα it was found that the central segment (50-89) of the protein has significant anti- HIV activity *(71)*. Interestingly the same 50-89 region of the protein is known as the metal binding region in Protα. This topic will be further discussed in detail in section 1.4 and 1.5.

Protα is shown to interact with mutant Huntington protein. Huntington disease (HD), a fatal neurodegenerative disorder, is caused by a lengthening of the polyglutamine tract in the huntington (Htt) protein *(74)*. Over expression of Protα remarkably reduced mHtt-induced cytotoxicity in both non-neuronal and neuronal cell models expressing N-terminal mHtt fragments, whereas knockdown of Protα expression in the cells enhanced mHtt caused cell death. Deletion of the central acidic domain of Protα abolished not only its interaction with mHtt but also its protective effect on mHtt-caused cytotoxicity *(74)*.

Interaction with binding partners

Protα is known to interact with REV of HIV-1 and CREB *(70, 75)*. Rev stands for Regulator of expression of virion proteins. It shuttles between the nucleus and the cytoplasm and harbors both

a Nuclear Localization Signal (NLS) and a Nuclear Export Signal (NES) *(76)*. The function of this protein is nuclear export of unspliced RNA of HIV-1. Protα is known to bind the NES domain of REV through N-terminal lysine residues. This interaction is shown to significantly increase in the presence of zinc *(75)*. Zinc bound Protα is partially folded resulting in enhanced interaction with REV *(60, 75)*.

HIV infected patients have zinc deficiency and decreased intracellular concentration of Protα *(77, 78)*. The fact that Zn^{2+} binding region of the protein falls in the anti-HIV domain of the protein (50-89 segment) and the enhancement of the interaction of Protα to the NES region of REV in the presence of Zn^{2+}, leads to the speculation that down regulation of zinc and Protα in HIV-1 infected cells is one of the viral defense mechanisms to release REV from bound Protα and carry out its function for HIV replication *(64, 79)*. An elucidation of the function of Protα and its Zn^{2+} binding characteristics opens new possibilities for the understanding of role of the protein in several biological pathways.

Protα interacts with a cellular transcription factor, CREB. CREB stands for cAMP response element binding protein. Direct evidence for physical interaction between Protα and CREB was obtained by glutathione pull down experiments. Protα interacts with the versatile co-activator of transcription CBP (CREB binding partner) both *in vivo* and *in vitro* and stimulates gene transcription. The site of interaction was mapped within the N-terminal domain of CBP (residues 1–771) and a region of Protα composed of two polyglutamic acid stretches. This binding could be due to the specific sequence characteristics of this acidic domain of Protα and/or its local conformation. The interaction of Protα with the CREB protein and its ability to stimulate transcription suggests an important role for this acidic polypeptide in gene expression *(70)*.

1.4 Metal binding characteristics of Prothymosin-α

Several investigations on the structural properties of Protα have been carried out by means of NMR, CD, SAXS, gel filtration, ANS fluorescene *(43, 60, 75)*. The studies have shown that at physiological pH the protein is in a random coil like conformation. However, the structure of Protα has been shown to be influenced by environmental factors like pH and counter ions *(59)*. At highly acidic pH Protα is shown to be transformed into a partially folded conformation known as pre-molten globule state *(80)*. Zn^{2+} binding of the protein at physiological pH has been also shown to induce a change in the secondary structure of Protα.

There is no literature information showing temperature effects on Protα. However, the effect of temperature on other natively unfolded proteins, like Protα, has shown that the proteins undergo a structural change adopting a partial α-helix conformation. The effect of temperature specifically on Protα will be discussed in Chapter 3.

Protα is known to bind several metal cations like Ca^{2+}, Mg^{2+}, Mn^{2+}, Zn^{2+}, etc. Among all these metal cations, only Zn^{2+} binding induces a structural change in the protein.

At physiological pH Protα is unfolded due to the strong electrostatic repulsion between acidic residues. It was believed that this repulsion could be reduced by the addition of oppositely charged ions *(43)*. To test this hypothesis several divalent metal cations like Ca^{2+}, Mg^{2+}, Mn^{2+}, Zn^{2+}, etc were added to the protein and using CD spectroscopy it was found that the protein undergoes a partial change in the secondary structure only upon Zn^{2+} addition (Figure 1.6A). A titration plot was constructed by plotting ellipticity versus wavelength which shows the saturation behavior (Figure 1.6B) and shows that the protein was saturated at 75 eq of $ZnCl_2$.

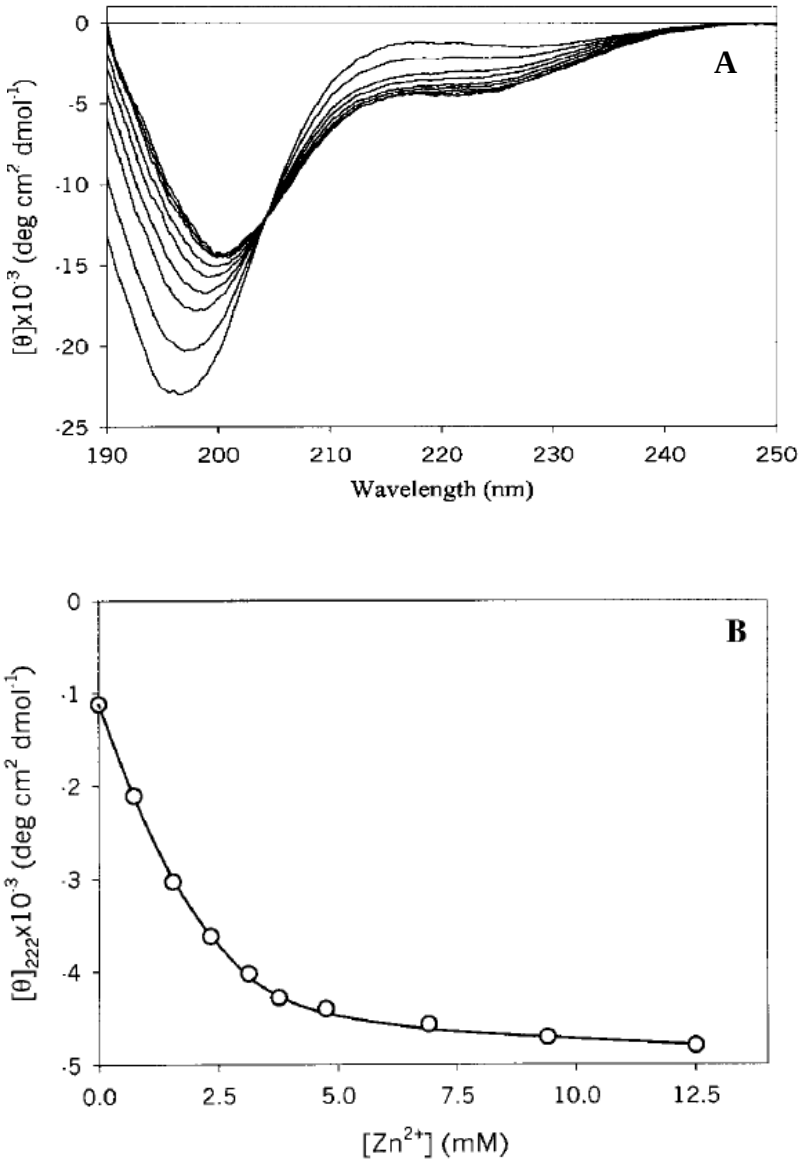


Figure 1.6: Zn^{2+}-induced changes in Protα far UV CD spectrum. A) Far UV CD spectrum of the full length Protα at different Zn^{2+} concentrations. B) The Zn^{2+} dependence of the $[\theta]_{222}$ value *(43)*.

However, this structural change occurred specifically in the presence of Zn^{2+}. The fact that no other positively charged divalent metal cation except Zn^{2+} could induce the structural change suggested that Zn^{2+} binding may be important for the protein's function.

The same study using SAXS looked at the radius of gyration (R_g) of native and metal bound Protα. The lower the R_g value the more structured the protein. The R_g value of the protein at low pH or with $ZnCl_2$ decreased reflecting considerable compaction of the protein. X-ray scattering also has shown that acidification or addition of Zn^{2+} to Protα induced a more ordered conformation.

Chichkova *et al.* has used equilibrium dialysis to show the binding of Protα to Zn^{2+} in the presence of other ionic competitors *(75)*. It was found that the binding is selective for Zn^{2+} and not just taking place due to electrostatic interactions. The parameters of zinc binding by Protα were determined using Zn^{2+} at different concentrations. Two buffer systems were used in this study, one with "low ionic strength" and another with "high ionic strength". The results show that under low ionic conditions Protα binds up to 13 zinc ions per one protein molecule with an apparent K_D = 40 μM. Under high ionic strength, Protα was found to bind 5 Zn^{2+} with an apparent K_D = 230 μM. The data suggests that Protα has several Zn^{2+} binding sites which differ in affinities based on the ionic conditions.

More info about the metal binding region of Protα have been revealed by chemical shift mapping by Yi *et al. (60)*. This study used NMR spectroscopy and ESI-MS to characterize the structure and dynamics of Protα and its interactions with zinc ions.

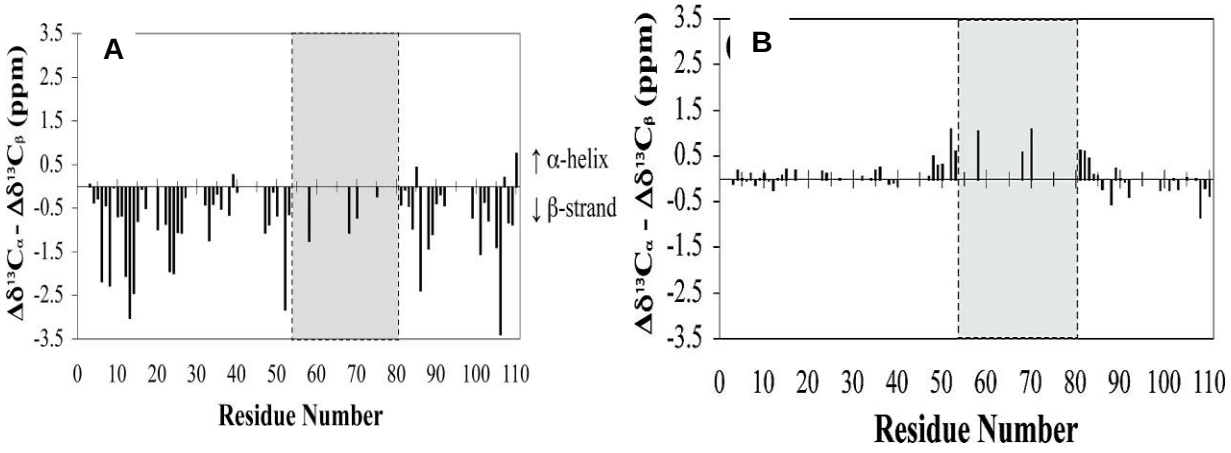


Figure 1.7: Reference-Independent[$\Delta\delta^{13}_{\alpha}C - \Delta\delta^{13}_{\beta}C(ppm)$] secondary chemical shifts of (A) Protα (B) Protα in the presence of 6 mM of Zn^{2+}. Positive and negative strands indicate the helical and β-strand propensities, respectively. The shaded box indicates the glutamic acid rich region, which contains 20 unassigned acidic amino acid residues *(60)*.

The backbone resonance assignments of $^{13}C/^{15}N$-labeled Protα were assigned by collecting a set of five heteronuclear triple resonance experiments, HNCACB, CBCA(CO)NH, CCCTOCSY, HNCACO, and HNCO *(81)*. For zinc titration experiments, ^{1}H-^{15}N HSQC spectra of 0.1 mM ^{15}N-labeled Protα were recorded in the presence of different concentrations of $ZnCl_2$ (0-3 mM). It has revealed that the protein binds Zn^{2+} via its C-terminal residues (48-110). The results from these studies indicate that Zn^{2+} binding to the central region have a significant influence on the dynamics of the target binding sites of Protα. The above Figure 1.7 represents reference independent chemical shifts of each residue in Protα before and after the addition of zinc. The central Glu rich region was left unassigned, because of the homogeneity in the sequence.

Pulsed field gradient experiments were performed to obtain the hydrodynamic radius of Protα. These experiments helped to study overall compactness of Protα upon Zn^{2+} binding. The

hydrodynamics radius, $R_{h,}$ of Zn^{2+} titrated Protα decreased compared to the R_h of native Protα. This once again confirms transition of Protα to more compact conformation upon metal binding.

Backbone dynamics of Protα were reexamined using relaxation measurements and steady state NOE experiments. In the absence of Zn^{2+} the measured relaxation rate (R_1) (defined as the rate at which the molecule relaxes) values of Protα fell in the range 1.2-2.0 s^{-1}. Figure 1.8 shows that upon addition of 6mM Zn^{2+} there was a significant increase in the R_1 values of residues 48-82 suggesting that it is the Zn^{2+} binding region.

The steady state ^{1}H-^{15}N NOE is sensitive to fast internal motions of the backbone, providing a good indicator of protein mobility on the picosecond-nanosecond time scale. For a folded protein with a molecular weight corresponding to that of Protα the NOE values with restricted mobility would be expected to be around 0.8 *(82)*. A lower NOE value indicates increased local flexibility. For native Protα all the NOE values were negative indicating a highly unstructured nature of the protein (Figure 1.9). With the addition of Zn^{2+} the NOE values became positive for residues 48-110 indicating a restricted mobility of the residues upon Zn^{2+} addition.

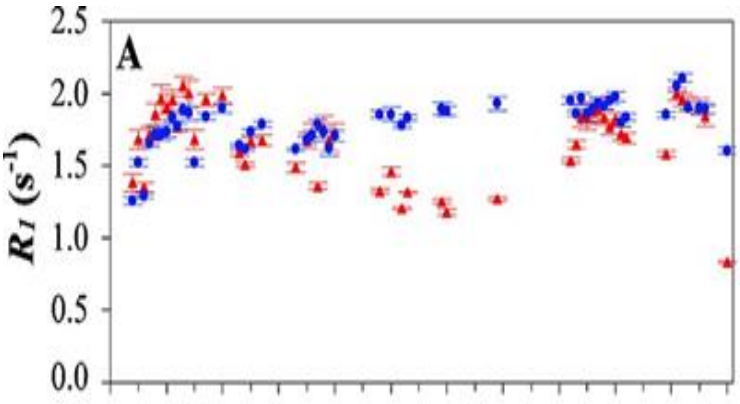


Figure 1.8: Backbone ^{15}N longitudinal relaxation measurements of Protα. The experiments were carried out in the presence (blue) and absence (red) of 6 mM $ZnCl_2$ *(60).*

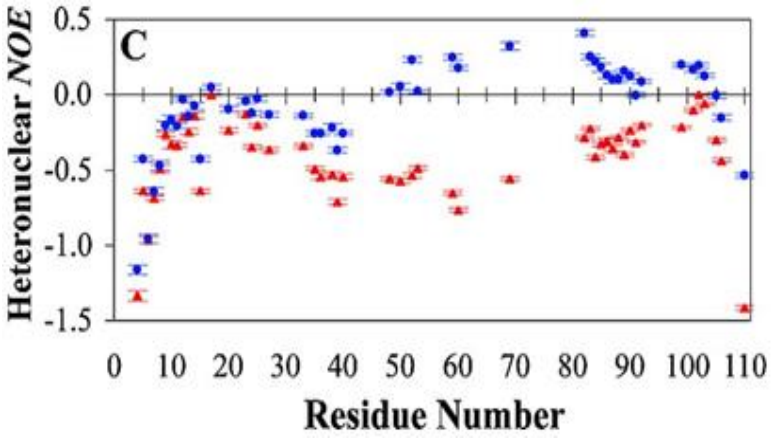


Figure 1.9: Backbone steady-state ^{1}H-^{15}N NOE of Protα. The experiments were carried out in the presence (blue) and absence (red) of 6 mM $ZnCl_2$ *(60).*

Influence of Zn^{2+} on binding of Protα with its partners

Protα has many binding partners. It interacts with REV of HIV-1 and Histone H1 *(83, 84)*. Rev stands for Regulator of expression of virion proteins. It shuttles between the nucleus and the cytoplasm and harbors both a Nuclear Localization Signal (NLS) and a Nuclear Export Signal (NES) *(76, 85)*. The function of this protein is nuclear export of unspliced RNA of HIV-1. Protα is known to bind the NES domain of REV through N-terminal lysine residues. The effect of Zn^{2+} on the binding of Protα to REV was observed *(76).* Both in the presence and absence of Zn^{2+}, Protα is shown to interact with REV. However, the interaction of Protα with REV was significantly enhanced in the presence of Zn^{2+} as determined by a 'pull-down' assay. ^{32}P-labeled Protα was incubated with Ig-G sepharose immobilized zz-rev and zz (1 µg of the corresponding protein per sample) in the binding buffer in the presence and absence of 10 µM Zn^{2+}. The bound protein was eluted by SDS containing buffer and analyzed by 12% PAGE followed by autoradiography. The K_D value is estimated to be 1 µM for zinc binding by the Rev+Protα complex.

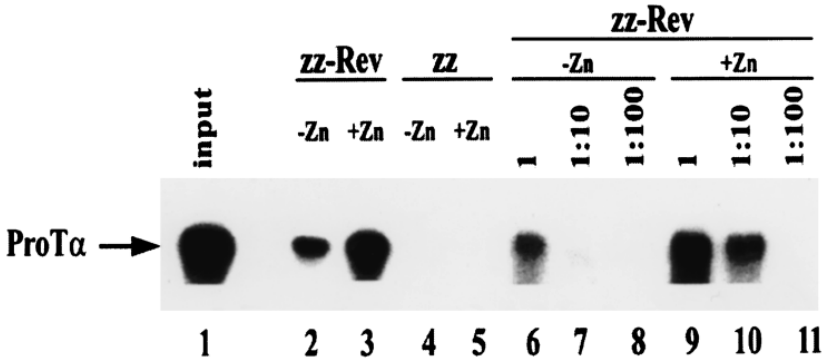


Figure 1.10: Protα binding to REV of HIV-1 is Zn^{2+} dependent *(74, 75, 3).* 'Input'(lane 1), the amount of [^{32}P]Protα added to each sample. 1:10 and 1:100 denote 10-fold and 100-fold dilution of the [^{32}P]Protα with unlabelled Protα lacking artificially added sequences prior to incubation.

The effect of divalent cations on the interaction of Protα with isolated NES region of Rev was also assayed, using a pull-down technique to understand the influence of metal cations. Both in the absence and presence of varying concentrations of Ca^{2+} and Mg^{2+} ions a barely detectable binding of Protα to NES was observed. However, in the presence of Zn^{2+} the interaction of Protα with NES of REV was enhanced. The NES+Protα complex formed in the presence of Zn^{2+} ions appeared to be unstable and could be dissociated easily.

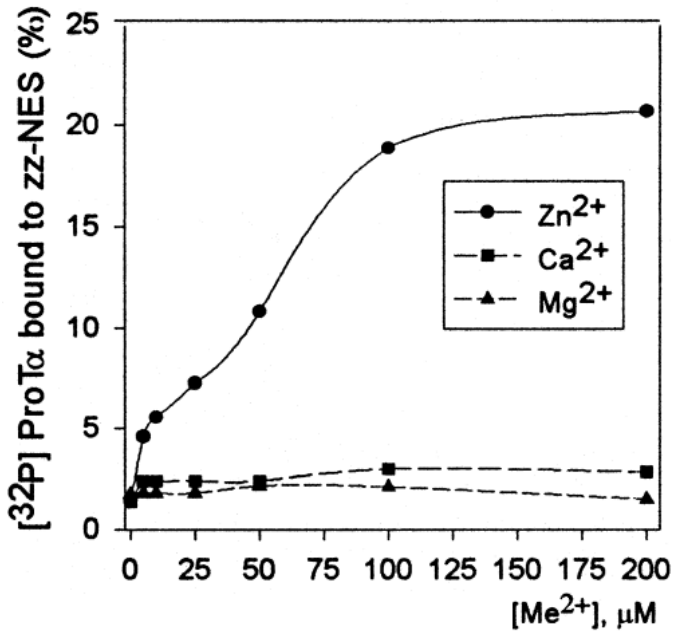


Figure 1.11: Zn^{2+} promotes Protα binding to the NES region of REV *(75)*

More characteristics of the Zn^{2+} and Protα+REV complex have come to light. HIV infected patients have zinc deficiency and decreased intracellular concentration of Protα *(77, 78)*. The fact that Protα binds NES region of REV, enhancing the interaction in the presence of zinc, leads to the speculation that down regulation of zinc and Protα in HIV-1 infected cells is one of the viral defense mechanisms to release REV from bound Protα and carry out its function for HIV replication *(64, 79)*. Thus, a detailed understanding of Zn^{2+} binding sites in Protα and further studying of binding of the Zn^{2+} to Protα+NES of REV is highly essential.

Protα does not share homology with any of the well characterized zinc binding proteins such as zinc fingers or superoxide dismutase. It does not have the most common zinc binding amino acids such as histidine and cysteine, which means that the Zn^{2+} binding motif is probably unique and might reveal an answer to how proteins that lack cysteine and histidine bind Zn^{2+}. It might also provide insight into how IDP's bind metals.

Section 1.5 Research goals and overview of research strategy

The goal of this project is to investigate metal binding in natively unfolded proteins, using Protα(50-89)N50W as a model peptide. The protein is unusual in that aspartic acid (Asp) and glutamic acid (Glu) make up approximately 50% of its 110 residues and it lacks typical metal coordinating residues such as histidine (His) and cysteine (Cys). However, it is shown to bind several metal cations (Mg^{2+}, Zn^{2+}, Ca^{2+}, etc.) and among all the metal cations only Zn^{2+} could induce a structural change in the protein *(43)*. Given the high homogeneity of the sequence, further study of Zn^{2+} binding using 2D nuclear magnetic resonance (NMR) studies (which is mostly used to study metal binding in proteins) has been a challenge. Due to extensive signal overlap NMR had failed to determine which Asp and Glu residues are directly involved in Zn^{2+} binding *(60)*. Further, Zn^{2+} is a spectroscopically silent metal. Our work using Protα(50-89)N50W should lay ground work for answering questions like: where is Zn^{2+} binding to Protα and how Zn^{2+} binding is affected by varying factors like temperature. We have also examined if Zn^{2+} binding is solely due to electrostatic interactions with the highly negatively charged sequences or if it is a sequence specific interaction.

As mentioned in section 1.1 the extent of information known about metal binding in natively unfolded proteins is limited. Understanding the Zn^{2+} binding in natively unfolded protein like Protα serves as a case study for how IDPs in general may bind metals so we may gain a mechanistic understanding of the specific structural and functional roles protein disorder plays with regard to metal binding.

Aim: To determine if zinc binding to prothymosin-alpha is specific - meaning that a unique set of residues provide the coordinating ligands or if many energetically "degenerate" sites exist within the central region (Figure 1.12)

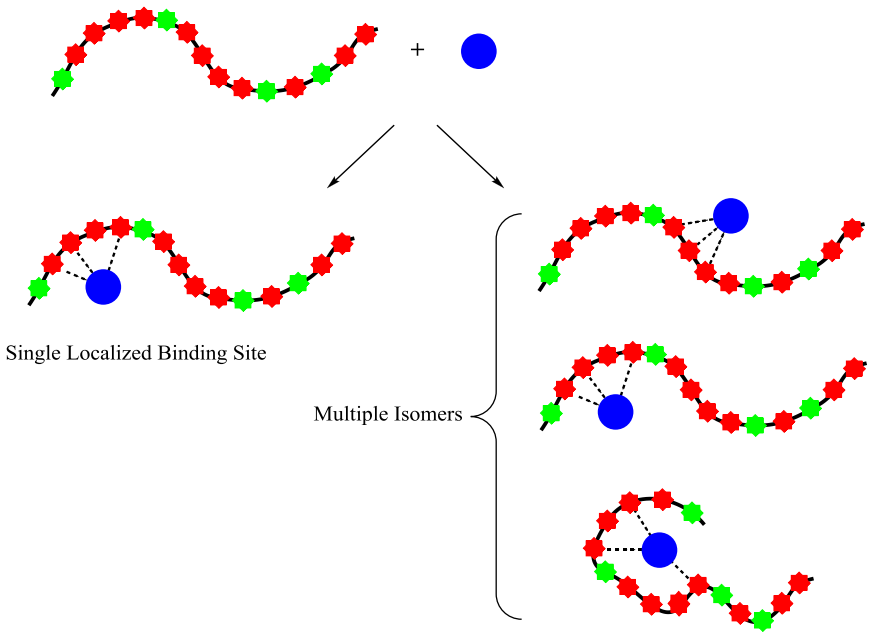


Figure 1.12: The protein sequence is displayed as red and green stars, where the red residues may serve as ligands, and the blue sphere represents Zn^{2+}. Two possibilities are envisioned where (1) a single defined set of red residues constitute the Zn^{2+} binding site or (2) numerous different combinations of the red residues may constitute a binding site leading to the formation of numerous isomers.

Specific goals:

1.1 Develop a synthetic method using 4-pyridinemethanol as a side chain protecting group on Asp and Glu to improve the synthesis, purification, and characterization of highly negatively charged sequence of Protα(50-89)N50W

Synthesis, purification, and characterization of peptides derived from acidic sequences like Protα has been challenging and requires development of a robust synthetic methodology. For all our

metal binding studies Protα derived peptides were chemically synthesized in order to keep them free of any biological extracts. Also, chemical synthesis of peptides is essential for ^{13}C labeling of peptides for NMR studies (discussed later in NMR studies subsection). So our first focus was to development a synthetic approach that allows us to improve the purification and characterization of acidic sequences like Protα(50-89)N50W.

The Zn^{2+} binding region of Protα has been narrowed down to the central 48-110 segment *(59).* However, the suspected metal binding residues in Protα, Glu and Asp are concentrated in the central 50-89 region which made us choose 50-89 as our model peptide for metal binding studies. Our synthetic work focuses on the central 50-89 segment of Protα and all the metal binding studies use the same segment.

Considering that the 50-89 sequence contains more than 50% of acidic residues; synthesis, purification, and characterization of the sequence is challenging. Our group has been working to improve the synthesis, purification, and characterization of the Protα(50-89) segment. During synthesis we encountered problems like difficult couplings. Because of the high negative charge, reverse phase chromatography and positive ion mode on ESI-MS cannot be used for the purification and characterization, respectively, of the synthesized peptides.

To overcome these problems a strong anion exchange (SAX) purification method was developed in our lab *(86)*. The method was successful in purification and characterization of Protα(50-89). However, the eluent used in SAX contains 1M NaCl which requires extensive dialysis (6 hrs) of the purified sample before injection onto MS for characterization. The main drawback of dialysis is it often dilutes the sample and sometimes results in sample loss making signal detection tough on ESI-MS. Benzyl ester protected aspartic and glutamic acid were used in the synthesis. The

group is acid resistant and is not lost upon treatment with acid. It masks the negative charge of aspartic and glutamic acid thus making characterization possibly on positive mode of ESI-MS. However, because of the lack of charge on benzyl ester it was found to have its own limitations (discussed in detail in section 2.2).

Here our goal was to develop a synthetic method for the synthesis of negatively charged sequences that makes purification and characterization straight forward. In this direction, 4-pyridinemethanol, which is structurally similar to benzyl alcohol, was used as a side chain protecting group on negatively charged Asp ad Glu residues. 4-pyridinemethanol is also acid resistant. The advantage of using 4-pyridinemethanol over benzyl ester is that the pyridine ring on 4-pyridinemethanol readily accepts a proton and flips the negative charge on carboxylic acids of Asp and Glu to a positive charge making the characterization of negatively charged sequences possible on positive ion mode of ESI-MS. Whereas the benzyl ester group flips the charge from negative to neutral thus making characterization of large peptides challenging using positive mode of ESI-MS. The new Asp and Glu derivatives of 4-pyridinemethanol were further applied in the synthesis of model peptides to test their efficiency.

1.2 Determine the structural changes in the central segment of Protα in the presence of environmental factors like Zn^{2+} and temperature

CD studies were carried out to determine the structural changes in the central 50-89 segment of Protα in the presence of Zn^{2+}. Protα is classified as a natively unfolded protein *(59)*. In the presence of Zn^{2+} it was found to undergo structural rearrangement *(43)*. The protein is known to bind 13 Zn^{2+} ions at low ionic conditions (K_D = 40 µM) and 5 Zn^{2+} ions at high ionic conditions (K_D = 230 µM) *(74)*. Although it is known that the central region of the protein is the metal binding region *(60)*, the portion of the protein that undergoes majority of the structural

rearrangement has not been determined. We hypothesized that the structural changes in the protein occur primarily in the central region. Our first aim was to look for the structural changes in the central segment (50-89 segment) of the protein in the presence of Zn^{2+} by using Circular Dichroism (*CD*). CD measurements were made by titrating zinc into the peptide. Spectra obtained from these experiments were compared to the CD spectrum of the full length protein *(43)*. Next, we studied temperature dependent structural change in Protα(50-89)N50W both in the presence and absence of Zn^{2+}. A polyglutamic acid sequence was used to study if Zn^{2+} binding in Protα is solely due to electrostatic interactions or if there is sequence specificity.

1.3 Investigate the structural changes in the Protα(50-89)N50W upon thermal denaturation using differential scanning calorimetry (DSC)

A DSC study on Protα(50-89)N50W was carried out to study the conformational stability of the peptide in the presence and absence of Zn^{2+}. The studies complement the data obtained from CD.

1.4 Further refine the map of the Zinc binding region of Protα

The central question addressed here is whether there is a single localized Zn^{2+} binding site or are there several degenerate Zn^{2+} binding sites in Protα.

An earlier study to map out Zn^{2+} binding region of Protα has shown that 2D NMR had difficulties to study Zn^{2+} binding on a residue level (58). The central segment of the protein due to high homogeneity is shown to have extreme signal overlap causing trouble to assign amino acid chemical shift in this region. Hence, a new approach is needed as regular NMR fails (signal overlap and too much homogeneity). We envision using ^{13}C labeled amino acids to study Zn^{2+} binding by making relaxation measurements. In order to prove this approach works, the way to do this is by picking a system we absolutely know binds Zn^{2+} and see if we can detect

sufficiently large changes in the T_1 values. This was achieved using a simple and inexpensive model complex of EDTA and Zn^{2+}. Later on the relaxation measurements will be performed on a series of ^{13}C labeled Protα peptides in the presence and absence of Zn^{2+}.

As a first step, we developed a ^{13}C NMR relaxation method to detect Zn^{2+} binding at the amino acid level using the metal-chelator EDTA as a model compound. This has laid down ground work for a core set of studies which will involve application of this method to series of Protα peptides spanning the metal binding region which contain site-specifically ^{13}C labeled residues. This system serves as a case study for how IDPs in general may bind metals so we may gain a mechanistic understanding of the specific structural and functional roles protein disorder plays with regard to metal binding.

We have chosen ethylenediaminetetraacetic acid (EDTA) as the model compound for the proof-of-principle ^{13}C relaxation studies (Figure 1.13) because of the structural similarities between EDTA and carboxylate groups of Glu and Asp. Also, it is known that EDTA is strong chelator of Zn (logK = 15.4) *(87)*. The spin-lattice (T_1) relaxation times of EDTA with and without Zn^{2+} were measured. In particular, we focused on the T_1 values for the carbonyl carbons as these are the closest spin-active nuclei to the Zn^{2+} ion and thus should be most sensitive to its presence or absence. To corroborate with the T_1 values obtained, the peptide studies will be carried out by adding the ^{67}Zn isotope to the sample to look for detectable line broadening in the NMR signal. The experimental details will be explained in section 3.2.

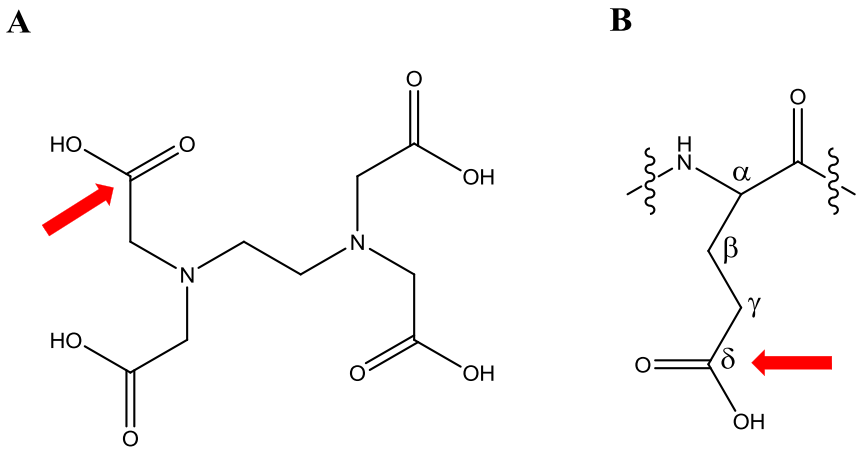


Figure 1.13: A) Structure of EDTA with one of its four carbonyl carbons indicated by red arrow, B) Structure of glutamic acid in the context of a peptide sequence. The carbonyl carbon of its carboxyl side chain is indicated with a red arrow.

So far it is estimated that over 30% of proteins catalogued in the protein data base require metal ions to perform their biological activities (1). In surveys of select eukaryotic genomes more than 30% of sequences are predicted to have disordered regions 50 residues or greater (12) and hundreds have been classified as intrinsically disordered using biophysical techniques such as CD and NMR spectroscopy (12). Thus, there is a great region of overlap here that remains to be explored. Ultimately, we seek to gain a mechanistic understanding of the specific structural and functional roles protein disorder plays with regard to metal binding. For instance, evidence is mounting that IDPs have faster kinetic binding profiles compared to their structured counterparts. The coupling of folding with binding by IDPs may lead to a significant reduction in binding free-energy barrier and thus an acceleration of the binding process (38). Further, Protα lacks typical metal binding residues such as His and Cys. Since a characteristic of many IDPs is high net charge this may stimulate the discovery of metal binding regions in proteins that would

otherwise go overlooked. Our case study of Protα, with key initial steps outlined here, serves as a means to gain insight into the type of issues mentioned above.

Each of these specific goals will be discussed in detail as a chapter.

Chapter 2

The 4-pyridylmethyl ester as a protecting group for glutamic and aspartic acids: 'flipping' peptide charge states for characterization by positive ion mode ESI-MS

Section 2.1 Introduction

The most common methods for the synthesis of peptides are solid-phase peptide synthesis (SPPS) and solution phase synthesis. Both the methods can be carried out in a linear fashion or by chemical ligation. In linear synthesis each amino acid can be added sequentially, and the synthesis can be carried out on an automated synthesizer or in a manual way. The chemical ligation method involves fragment condensation of peptide fragments to form the target sequence.

Synthesis, purification and characterization of highly negatively charged sequences have always been a challenge *(86)*. Often the standard synthesis method has to be modified to achieve the synthesis of the target acidic sequence. A standard linear synthesis requires 20-120 min for the coupling of each amino acid depending on the choice of chemistry and synthetic protocol followed. However, with acidic sequences the coupling times have to be doubled or tripled to achieve acceptable acylation yields *(88)*. Even after careful implementation of these methods the long coupling times of the target sequence allow kinetically slower side chain reactions to become competitive thus leading to the failure of the acidic sequence synthesis *(89, 90)*. A serious problem affecting the fragment condensation method in acidic sequences is the decreasing solubility of peptide fragments in organic solvents making the ligation tricky.

Protα is an acidic protein. Chemically synthesized Protα is more advantageous to use than the cell extracted protein for our studies with this protein as the chemically synthesized peptide can

be modified at any position in the sequence during the synthesis and it is free of unwanted biological components that cell extracted Protα might have.

Majority of the negative charge is located in the central 50-89 region of Protα which is found to be important for its activity (Figure 1.5). The 50-89 segment is found to be essential for metal binding *(58)* and anti-HIV activity *(64, 64, 71-73)*. So, earlier work carried out in our lab on Protα focused on the central 50-89 segment.

Synthesis:

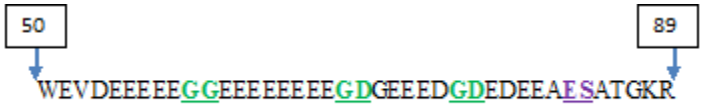
Several peptides derived from Protα, mainly focused on the 50-89 region were synthesized by linear solid phase synthesis on an automated Protein Technologies Inc. PS3 synthesizer or a manual CEM microwave synthesizer.

1.Linear Solid Phase Peptide Synthesis (PS3 Synthesizer):

Synthesis of Protα (50-89) N50W was attempted several times on a PS3 synthesizer. Due to many unsuccessful attempts a few modifications were made to the synthesis to make a few couplings proceed with higher efficiency.

The difficult steps were identified as couplings of glycine to aspartic acid, and glycine to glycine. In particular, the coupling of Gly59 to Gly60, Gly69 to Asp70, Gly76 to Asp77 were found to be challenging (underlined). To enhance the coupling efficiency at these positions, dipeptides Fmoc-Gly-(Dmb)Gly-OH, Fmoc-Asp(otBu)-(Dmb)Gly-OH were at their respective positions and residues 83 and 84 were substituted with a dimethyloxazolidine (pseduoproline) dipeptide. Pseudoproline helps to prevent aggregation early on in the synthesis. The dipeptide prevents secondary structure formation which often prevents amino acid couplings from occurring.

Placement of the pseudoproline earlier in the sequence (towards C-terminus) prevents the structure formation towards the N-terminus. Note the direction of SPPS is from C to N.



Sequence 2.1: Protα (50-89) N50W showing difficult coupling positions in green. The pseudoproline ES dipeptide unit is shown in purple.

Other modifications include doubling the coupling time (double couplings), changing the coupling reagent to HATU instead of the standard HBTU reagent, and using a low loading resin. The low loading resin was expected to prevent aggregation between nearby growing peptide strands, thus allowing the coupling reactions to go to completion. These new strategies were initially used individually (for example increase the coupling times) and later on in combinations (for example increase coupling times and also substituted HATU instead of HBTU), in all the syntheses. However, we were not successful in any of our trials.

2. Microwave Synthesis**:**

Microwave synthesis is a strategy for protein synthesis which has the potential to yield very pure proteins. Protα(50-89)N50W was synthesized successfully on a CEM Discover microwave synthesizer *(91)*. Two distinct points about microwave synthesizer are: The bulk sample was heated to 75 ^{o}C and the coupling and deprotection times were reduced to five and three minutes, respectively. SPPS is performed at room temperature whereas microwave synthesis is performed at elevated temperatures. The higher temperatures in microwave synthesizer promotes the

disruption of secondary structure (there is more energy to overcome intramolecular H-bonding interactions, etc.) and the reactions proceed faster at higher temperature.

Purification:

Due to the high number of hydrophilic residues in the sequence, retention on the reverse phase HPLC column is poor. So the peptides synthesized on the automated and microwave synthesizers were purified using a strong anion exchange (SAX) column. The disadvantage to the SAX method is that, the elution buffer contains 20 mM glycine with 1M NaCl. The samples collected by this method contain salt and needs to be extensively dialyzed in order to characterize it on ESI-MS. Dialysis of the sample against 18 mΩ water resulted in formation of Na/TFA adducts that appeared as peaks spaced at 136 Da and dominate the mass spectrum (Figure 2.1) *(86)*. Water was replaced by ammonium acetate and formate in dialysis to avoid the Na/TFA adducts formation.

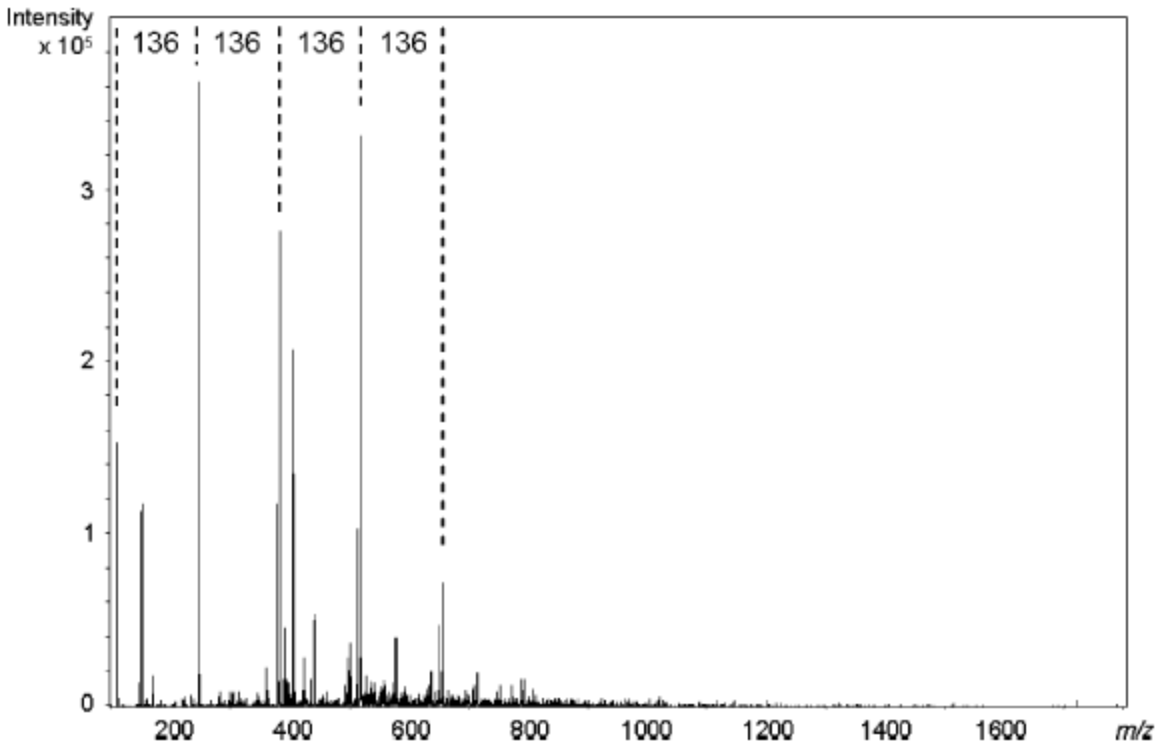


Figure 2.1: ESI-MS of Protα sample after strong anion exchange HPLC and dialysis against water. Na/TFA complexes, evidenced by peaks spaced at 136 Da, dominate the signal *(86)*.

However, dialysis is highly time consuming process and results in dilution of the sample. A synthetic strategy using benzyl ester as a side chain protecting group was developed to avoid the need for SAX *(92)*. Protα(50-105)N50W was synthesized in a linear fashion using benzyl ester protected aspartic and glutamic acids.

3. Linear synthesis using the benzyl ester (OBzl) protecting group:

In this method, selected glutamic acid residues (E) side chain were protected with benzyl ester (OBzl) instead of the standard tertiary butyl (tBu) group that was used in the previous two methods (sequence 2.2). The OBzl group is stable to acidic conditions and will not be removed

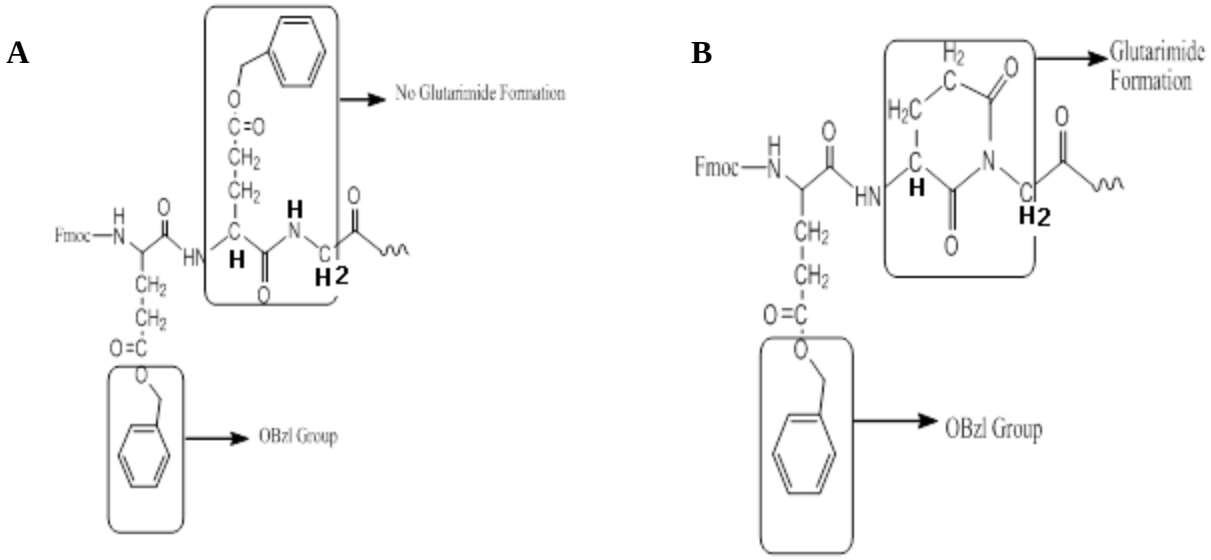
when resin and other side chain protecting groups are treated with trifluoroacetic acid. Thus it imparts neutral charge and hydrophobicity to the carboxylic acid side chain of Glu and aids in the purification using reverse phase C18 column and avoids the dialysis step. The flip of charge from negative to neutral facilitates characterization of the synthesized peptides on positive mode of ESI-MS. Later, the OBzl groups can be removed quantitatively by hydrogenation. Thus the purification and characterization steps were made easy. Out of the 33 negative charges in the sequence the synthesis required netralizing only 12 glutamic and aspartic acid residues to make the net charge of the sequence positive.

WEVDEEEEEGGEEEEEEEEGDGEEEDGDEDEEAESATGKRAAEDDEDDDVDTKKQK

Sequence 2.2: Protα (50-105) N50W sequence showing the benzyl ester protected glutamic acid residues in green.

The above strategy was applied in synthesizing Protα 50-105 (55mer) sequence. During the synthesis the coupling of Fmoc-Gly-OH to Fmoc-Glu-(OBzl)-OH was found to be problematic. We observed the formation of glutarimide (Figure: 2.2A and 2.2B) between these two amino acids. This problem was overcome by using Fmoc-Glu-(tBu)-OH for all the glutamic acids that precede glycine.

The synthesis was successful until step number 59, corresponding to a 47mer. As said, the benzyl ester protecting group imparts hydrophobicity to the peptide thus helped in the purification using reverse phase column and characterization on positive ion mode. However, with increasing length of the target sequence when the last 8 amino acids were added, the sequence could not be characterized on the positive ion mode because of excessive charge neutralization of the peptide by the benzyl ester group.



Weight of Glu(OBzl)- 219.59 Da

Weight of cyclized Glu(OBzl)-111.59 Da

Figure 2.2: A. Growing peptide chain showing no glutarimide formation **B**. Glutarimide formation between glycine and glutamic acid in a growing peptide chain. Note, glutaramide formation can be detected by a characteristic change is mass.

The limitations of the strategies used to synthesize Protα(50-89)N50W and Protα(50-105)N50W highlight the need for a protecting group that can aid in the purification and characterization of longer negatively charged sequences by positive ion mode MS.

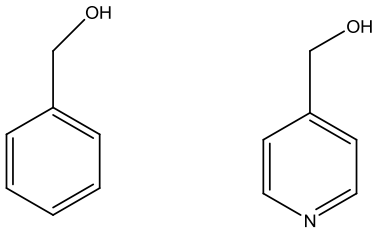
Section 2.2 Research Strategy

Like mentioned earlier, synthesis, purification, and characterization of highly acidic sequences has been challenging *(86, 93)*. Though the methods described in section 2.1 were successful in synthesizing the acidic Protα(50-89)N50W segment it would be more advantageous to avoid SAX for purification. To this end the benzyl ester protecting group was employed in the synthesis. However, this group also has its own disadvantage that was described in section 2.1.

Our goal was to develop a protecting group strategy that would allow purification by reverse-phase HPLC and characterization by ESI-MS in positive-ion mode. The 4-pyridylmethyl ester (abbreviated as OMePyr), also referred to as the 4-picolyl ester, introduced by Young possesses the necessary characteristics to make this possible; it is weakly basic yet non-nucleophilic and compatible with all the reaction conditions encountered in standard Fmoc- and t-Boc-based solid-phase peptide syntheses *(94, 95, 96)*. It was originally designed as a 'handle' that allowed for the solution-phase synthesis of peptides; placement at the C-terminus of a growing peptide chain enabled the peptide to be captured by an ion-exchange resin as excess reagents from a coupling or deprotection step were removed. The impetus for the development of the 4-pyridylmethyl ester stemmed from concerns about the inherent steric hindrance of the solid-phase methodology being pioneered by Merrifield. As convergent peptide synthesis became viewed as the prime strategy for synthesizing large peptides or proteins the generally low solubility of protected peptide segments was noted as a problem. The 4-pyridylmethyl ester protecting group was shown to increase the solubility of protected peptide segments allowing for purification by reverse-phase HPLC *(97, 98)* .

Comparison of 4-pyridylmethyl ester with benzyl ester

Esters of both these molecules are acid resistant and will not be removed when resin and other side chain protecting groups are cleaved from the peptide. So using any of these two groups as a side chain protecting group on negatively charged amino acids would keep those amino acids protected when the rest of the side chain protecting groups, and resin are cleaved from the peptide using strong acid like trifluoroacetic acid. In this way, we keep the negative charges on those amino acids masked and the groups impart hydrophobicity facilitating the purification of negatively charged sequences on reverse phase C18 column, and characterization on the positive ion mode of ESI-MS. Further, the groups can be removed by hydrogenation in the presence of palladium catalyst.



Benzyl alcohol 4-pyridinemethanol

Figure 2.3: Structures of benzyl alcohol and 4-pyridinemethanol

Selection of the 4-pyridylmethyl ester moiety over benzyl ester as a protecting group stemmed from the basic yet non-nulceophilic nature of the pyridine ring. The pK_a of the 4-pyridylmethyl ester in its acidic form should be close to that of pyridine, which is 5.25. This ensures that a significant proportion of 4-pyridylmethyl ester protected side-chains will be protonated under the conditions typically employed in RP HPLC and ESI-MS thus, the group flips the negative charge on carboxylic acids of aspartic and glutamic acid to a positive charge making the characterization

of negatively charged sequences possible on positive ion mode of ESI-MS. Whereas the benzyl ester group cannot be protonated and flips the charge from negative to neutral thus making characterization of large peptides challenging using positive mode of ESI-MS.

4-pyridinemethanol protected glutamic acid and aspartic acid are not commercially available. Different synthetic strategies were explored for the synthesis of 4-pyridylmethyl ester side chain protected glutamic and aspartic acid. N-terminally Fmoc protected and C-terminally tertiary butyl group (tBu) protected glutamic and aspartic acids were used as starting materials. The side chain carboxyl group was not protected and readily reacts with the incoming alcohol. After synthesis the 4-pyridylesters were purified and characterized.

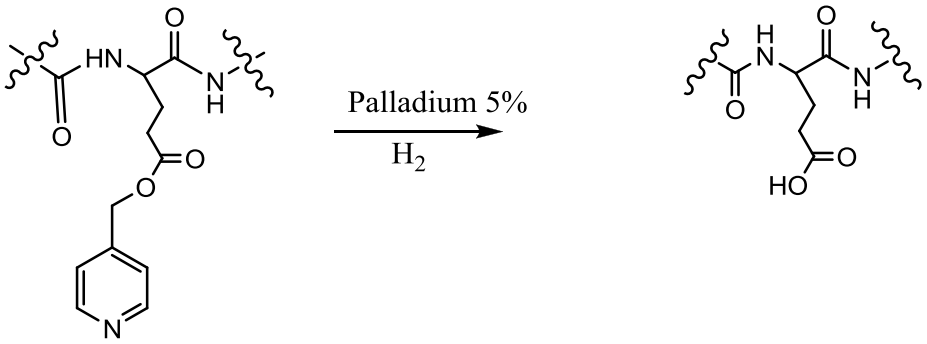
Later on a model peptide for testing the use of 4-pyridinemethanol-protected amino acids were used. We chose a model peptide that is easy to synthesize (sequence 2.3)

A) WGA**E**LAG B) WGA**D**LAG

Sequence 2.3: Model peptides showing the position of 4-pyridylmethyl-protected glutamic acid and aspartic acid in green

The highlighted residue E in the sequence marks the site of the proposed 4-pyridinemethanol-protected glutamic acid. As already said, the protecting group is acid resistant, so we expect it to remain intact when the resin and other side chain protecting groups are cleaved using strong acid. The purification can be carried out on a reverse phase column and characterization on positive ion mode of MS.

After purification and characterization of the peptide model was complete the 4-pyridinemethanol protecting group was removed by hydrogenation in the presence of palladium catalyst as shown in the scheme 2.1.



Scheme 2.1: Hydrogenation of 4-pyridinemethanol protected glutamic acid.

If the small peptide is successfully prepared, the new side chain-protected amino acid can then be applied in the synthesis of Protα peptide derivatives.

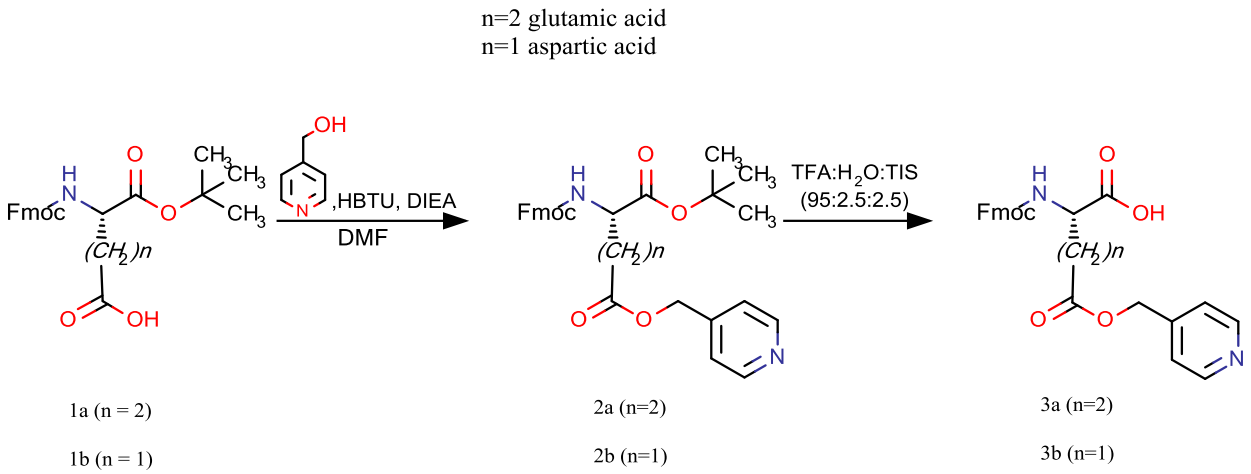
WEVDEEEEEGGEEEEEEEEEGDGEEEDGDEDEEAESATGKR

Sequence 2.4: Protα(50-89)N50W showing negatively charged residues in red and positively charged residues in green.

There are a total of 25 negatively charged and 2 positively charged residues in the sequence of Protα50-89N50W (sequence 2.4). By installing 4-pyridinemethanol aspartic or glutamic acid only in 5 positions in the sequence and continuing to use benzyl ester protected aspartic and glutamic acid in rest of the positions the overall charge on the peptide can be made positive. This minimizes the number of 4-pyridylmethyl esters of aspartic and glutamic acid required in the synthesis.

2.3 Synthesis, purification, and characterization of 4-pyridylmethyl ester of aspartic and glutamic acid.

Compound **3** was synthesized by straightforward esterification through *in situ* activation of **1** using HBTU and reaction with 4-pyridinemethanol followed by removal of the t-butyl protecting group (Scheme 2.2).



Scheme 2.2: Synthesis of Nα-Fmoc-L-glutamic acid γ-4-pyridylmethyl ester (**3a**) and Nα-Fmoc-L-aspartic acid γ-4-pyridylmethyl ester (**3b**) *(163).*

The starting material, intermediates, and final products were dissolved in acetone-d6 to obtain ^{1}H NMR spectra. ESI-MS was also obtained using 1:1 ratio of acetonitrile and water as solvent.

Synthesis of Nα-Fmoc-L-glutamic acid α-t.-butyl ester γ-4-pyridylmethyl ester (2a)

Compound **2** was synthesized by treating 0.71 mmol of N α-Fmoc-L-glutamic acid α-t.-butyl ester (**1**) with 0.71 mmol of HBTU, 0.71 mmol HOBt, 1.4 mmol 4-pyridinemethanol, and 2.85 mmol DIEA in 10 mL of DMF for 6 hr at room temperature. The solvent was removed under reduced pressure and the reaction mixture was dissolved in dichloromethane, washed with water,

dried over Na_2SO_4, filtered, and evaporated to give compound **2**. A light yellow solid product was obtained in 70.4% yield.

Characterization:

^{1}H NMR (acetone-d6) δ 8.69 (broad d, 2H), 7.87 (d, 2H), 7.81 (d, 2H), 7.70 (broad d, 2H), 7.44 (t, 2H), 7.35 (t, 2H), 6.07 (broad s, 1H), 5.33 (d, 2H), 4.36 (m, 2H), 4.25 (t, 1H), 4.13 (m, 1H), 2.57 (t, 2H), 2.30 (m, 2H), 1.45 (s, 9H).

ESI MS: *m/z* for $C_{30}H_{32}N_2O_6$ calc. 516.2260, $[M+H]^+$ found 517.2376.

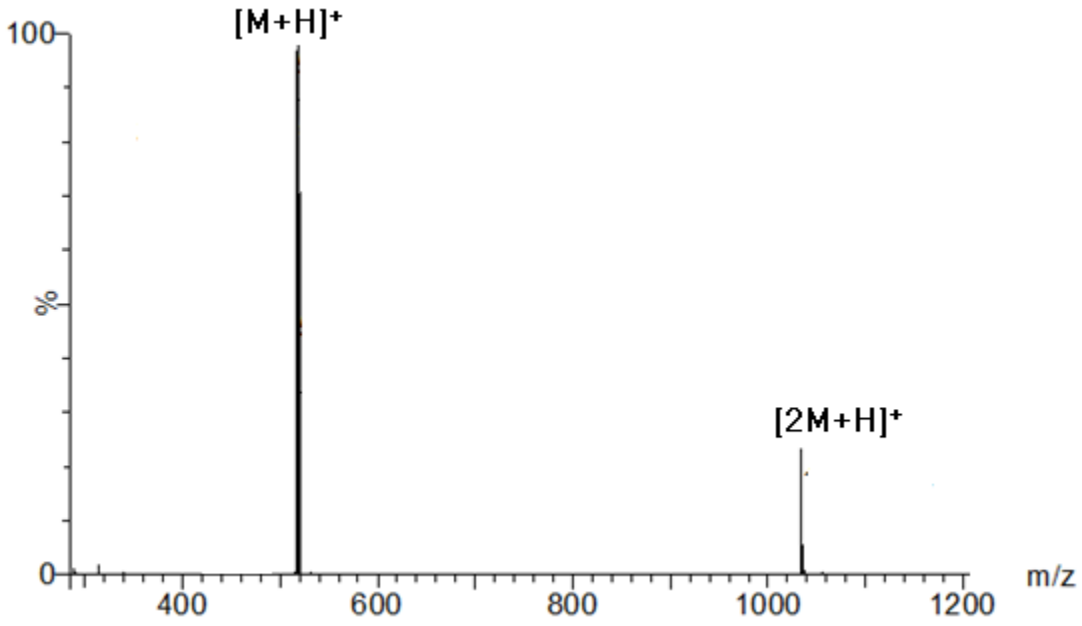


Figure 2.4: ESI-MS spectrum of Nα-Fmoc-L-glutamic acid α-t.-butyl ester γ-4-pyridylmethyl ester (**2a**) showing $[M+H]^+$ and $[2M+H]^+$.

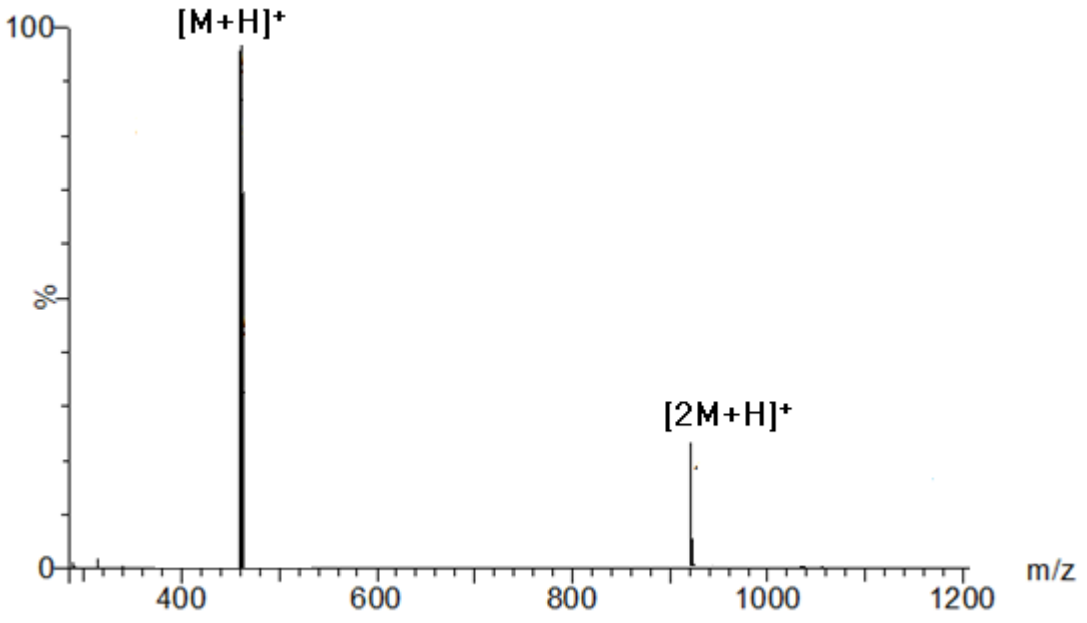
Nα-Fmoc-L-glutamic acid γ-4-pyridylmethyl ester (3a)

A 0.42 mmol quantity of compound **2** was dissolved in 1 mL of TFA/H_2O/TIS (95/2.5/2.5) and stirred for 3 hr. The product was precipitated with cold diethyl ether, centrifuged to collect a solid yellow product in 69.2% overall yield (i.e. calculated starting from compound **1**).

Characterization:

^{1}H NMR (acetone-d6) δ 8.69 (broad d, 2H), 7.87 (d, 2H), 7.81 (d, 2H), 7.70 (broad d, 2H), 7.44 (t, 2H), 7.35 (t, 2H), 6.07 (broad s, 1H), 5.33 (d, 2H), 4.36 (m, 2H), 4.25 (t, 1H), 4.13 (m, 1H), 2.57 (t, 2H), 2.30 (m, 2H).

ESI MS: *m/z* for $C_{26}H_{24}N_2O_6$



calc. 460.1634, $[M+H]^+$ found 461.1066.

Figure 2.5: ESI-MS spectrum of Nα-Fmoc-L-glutamic acid γ-4-pyridylmethyl ester (**3a**) showing $[M+H]^+$ and $[2M+H]^+$.

Nα-Fmoc-L-aspartic acid α-t-butyl ester α-4-pyridylmethyl ester (2b)

A 1.26 mmol quantity of Nα-Fmoc-L-aspartic acid α-t-butyl ester **(1b)** was treated with 1.26 mmol of HBTU, 1.26 mmol HOBt, 2.52 mmol 4-pyridinemethanol, and 5.04 mmol DIEA in 10 mL of DMF for 6 h at room temperature. The solvent was removed under reduced pressure and

the reaction mixture was dissolved in dichloromethane, washed with water, dried over Na_2SO_4, filtered, and evaporated to give compound **2b**. A light yellow solid product was obtained in 72.4% yield.

Characterization:

^{1}H NMR (acetone-d_6) δ 8.66 (broad d, 2H), 7.85 (d, 2H), 7.78 (d, 2H), 7.69 (broad d, 2H), 7.49 (t, 2H), 7.35 (t, 2H), 6.07 (broad s, 1H), 5.31 (d, 2H), 5.07 (t, 1H), 4.34 (m, 2H), 4.13 (m, 1H), 2.87 (m, 1H), 2.50 (m, 1H), 1.45 (s, 9H).

ESI MS: *m/z* for $C_{29}H_{30}N_2O_6$ calc. 502.2104, $[M+H]^+$ found 503.2776.

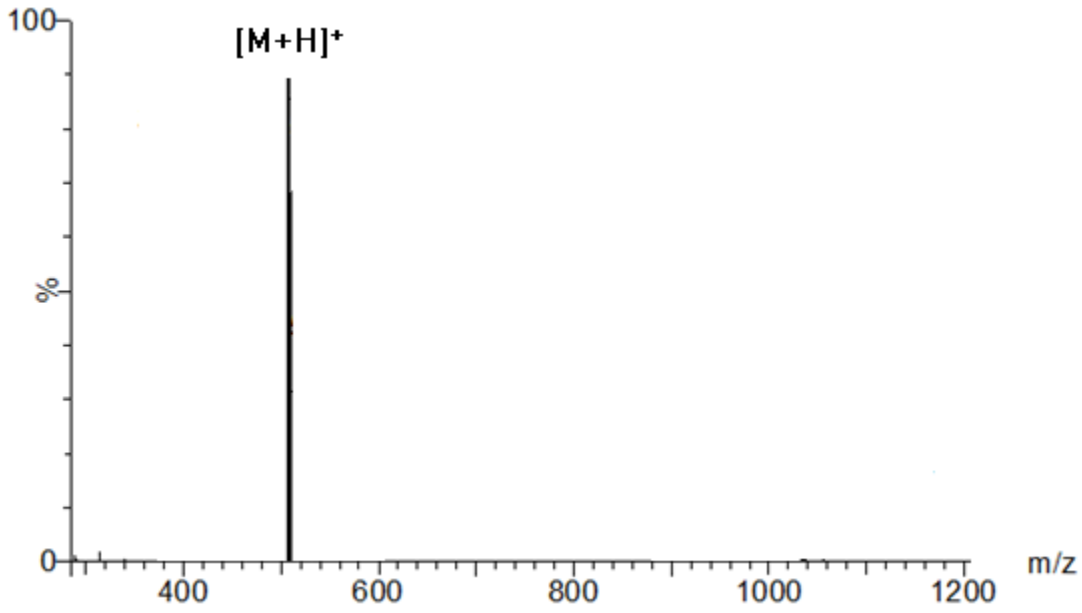


Figure 2.6: ESI-MS spectrum of Nα-Fmoc-L-aspartic acid α-t-butyl ester α-4-pyridylmethyl ester (**2b**) showing $[M+H]^+$.

Nα-Fmoc-L-aspartic acid α-4-pyridylmethyl ester (3b)

A 0.42 mmol quantity of Nα-Fmoc-L-aspartic acid α-t-butyl ester α-4-pyridylmethyl ester **(2b)**

was dissolved in 1 mL of TFA/H_2O/TIS (95/2.5/2.5) and stirred for 3 h. The product was precipitated with cold diethyl ether, centrifuged to collect a solid yellow product in 70.3% overall yield.

Characterization:

^{1}H NMR (acetone-d_6) δ 8.66 (broad d, 2H), 7.85 (d, 2H), 7.78 (d, 2H), 7.69 (broad d, 2H), 7.49 (t, 2H), 7.35 (t, 2H), 6.07 (broad s, 1H), 5.31 (d, 2H), 5.07 (t, 1H), 4.35 (m, 2H), 4.13 (m, 1H), 2.87 (m, 1H), 2.50 (m, 1H).

ESI MS: *m/z* for $C_{25}H_{22}N_2O_6$ calc. 446.1478, $[M+H]^+$ found 447.2066.

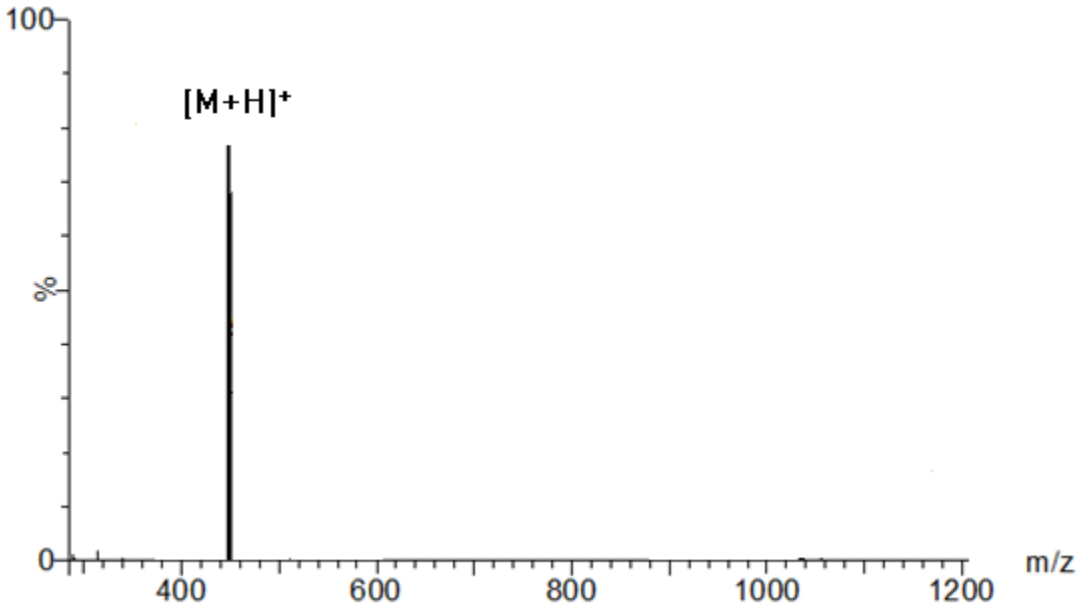


Figure 2.7: ESI-MS spectrum of Nα-Fmoc-L-aspartic acid α-4-pyridylmethyl ester (**2b**) showing $[M+H]^+$.

2.4 Application of 4-pyridylemethyl esters of aspartic and glutamic acid in the synthesis of model peptides

The ultimate goal of developing a protecting group is to be able to use it in peptide synthesis. The 4-pyridylmethyl ester protected glutamic acid (**3a**) and aspartic acid derivatives **(3b)** readily couple with free amines under the typical conditions employed for solid-phase peptide synthesis and were used in the synthesis of the heptapeptides Ac-WGAELAG-NH_2 (**5a**) and Ac-WGADLAG-NH_2 (**5b)** respectively. A hexapeptide Ac-WGAELAG-NH_2 with standard tBu protected glutamic acid was also synthesized (Scheme 2.3).

Peptide Synthesis

Peptides were synthesized using a Protein Technologies Inc. PS3 peptide synthesizer. All the peptides were synthesized on the 0.1 mmol scale using rink amide MBHA resin with 0.3 mmol/g substitution. Upon cleavage, the peptide produced from this resin was *C*-terminally amidated. Coupling of each amino acid was carried out in DMF using 4 equiv. Fmoc-amino acid/HBTU/HOBt and 4 equiv. DIEA with a 1 h coupling time. When compound **3** was incorporated into a peptide, 2 equiv. of Fmoc-amino acid/HBTU/HOBt were used. The Fmoc group was removed using 20% v/v piperidine in DMF with a reaction time of 30 min. The final peptide was acetylated using 2 mL of acetic anhydride and 4 eq. DIEA in DMF. After synthesis the resin was washed with dichloromethane, methanol and then dried. The dried resin was treated with a TFA/TIS/H_2O (95/2.5/2.5) solution for 2 h. The reaction mixture was filtered and the peptide precipitated with cold diethyl ether. The precipitate was collected, dissolved in water and freeze dried. The peptide was purified and characterized by HPLC and ESI-MS respectively and mass of the dried peptide was obtained to calculate percentage yields.

Purification by HPLC

Crude peptides were purified on a RP column using 0.1% TFA in water (Buffer A) and 0.1% TFA in acetonitrile (Buffer B). After injection, isocratic elution at 100% Buffer A was run for 3 min. followed by gradient elution from 100% Buffer A to 0% Buffer A in 37 min. with a flow rate of 1.0 mL/min. Peptide signals were detected at both 214 nm and 280 nm.

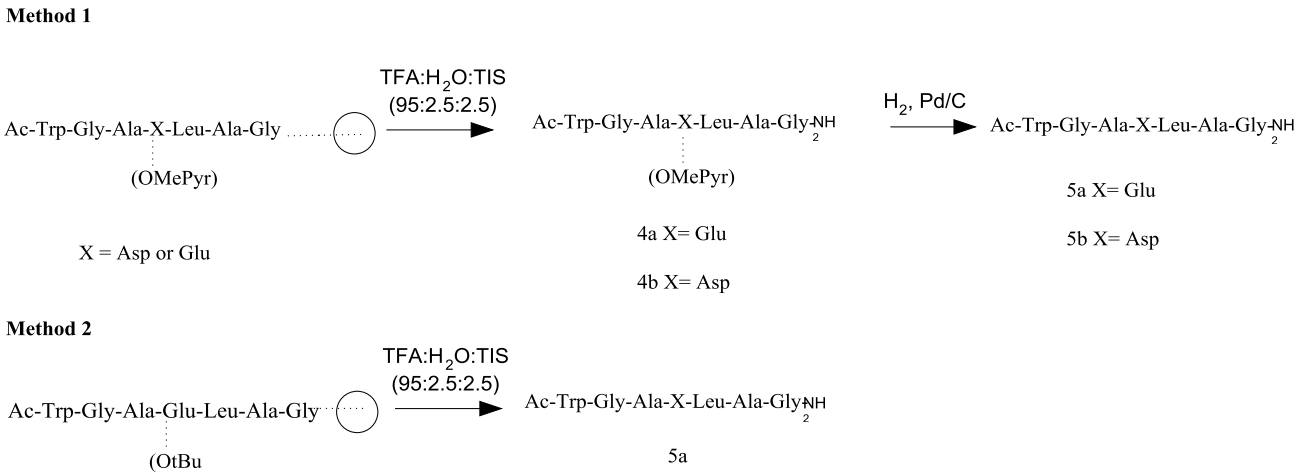
Characterization

All peptides were characterized by ESI-MS after elution from HPLC, and thus were dissolved in water/acetonitrile containing 0.1% TFA. Sample was delivered to the ESI source with a KD Scientific syringe pump at 10 µl/min, and the electric potential used to initiate ESI (i.e. capillary voltage) was 2500 V. The source temperature was set at 90 °C and the desolvation temperature at 275 ^{o}C with a cone gas flow of 10 L/hr and a desolvation gas flow of 500 L/hr.

ESI-MS: m/z for peptide **4a** calc. 834.3680, [M+H]+ found 835.3341; peptide **4b** calc. 820.4345, [M+H]+ found 821.4673; peptide **5a** calc. 743.4531, $[M-H]^-$ found 742.4236; peptide **5b** calc. 729.3367, $[M-H]^-$ found 728.3865.

Hydrogenation

As said in section 2.2 the 4-pyridylmethyl group was removed by overnight hydrogenation. The peptide was dissolved in 10 mL of ethyl acetate and 2 mg of 5% Pd/C was added. The reaction flask was placed under a hydrogen atmosphere of 70 psi and shaken overnight using a Parr hydrogenator. The Pd/C catalyst was removed by filtration through a fritted glass filter and the ethyl acetate was removed under reduced pressure.



Scheme 2.3: Synthesis of Ac-WGAELAG-NH_2 (**5a**) and Ac-WGADLAG-NH_2 (**5b**) using 4-pyridylmethyl ester side chain protection for glutamic acid and aspartic acid (method 1) and using t-butyl ester side-chain protection for gluatmic acid (method 2). The 4-pyridylmethyl ester group is abbreviated as –OMePyr and the circle represents the solid phase support *(163)*.

Percent Yields

The coupling efficiency of **3a** appears comparable with other standard side chain-protected Glu amino acids as peptide **4a** was obtained in 89.2% yield. The 4-pyridylmethyl ester was removed in near quantitative yield by hydrogenation using 5% Pd/C in ethyl acetate, thus exposing the carboxyl side chain of Glu (95.7% yield). Addition of the extra hydrogenation step decreases the overall yield of peptide 5a slightly (86.6% yield, Scheme 2, method 1) when compared with synthesis employing a standard t-butyl ester-protected Glu (88.9% yield, Scheme 2, method 2). An analogous heptapeptide Ac-WGADLAG-NH_2 **(5b)** was synthesized containing a 4-pyrdinemethyl ester-protected aspartic acid residue (Scheme 2, method 1). The peptide **5b** was obtained in 90.2% overall yield.

Discussion:

The peptides Ac-WGAE(OMePyr)LAG-NH_2 (**4a**) and Ac-WGAD(OMePyr)LAG-NH_2 (**4b**) were readily detected by positive mode ESI-MS as a singly charged species (Figure 2.8A). However, the peptides Ac-WGAELAG-NH_2 (**5a**) and Ac-WGADLAG-NH_2 (**5b**), which lack the 4-pyridylmethyl ester protecting group, were not observed by positive mode ESI-MS using common conditions and operating parameters (data not shown). This result is not surprising as they lack even weakly basic sites and are predicted to have an overall charge of zero in the 0-3 pH range. Peptides **5a** and **5b** were detected by negative ion mode ESI-MS as a singly charged species thus allowing for characterization in this case (Figure 2.8B).

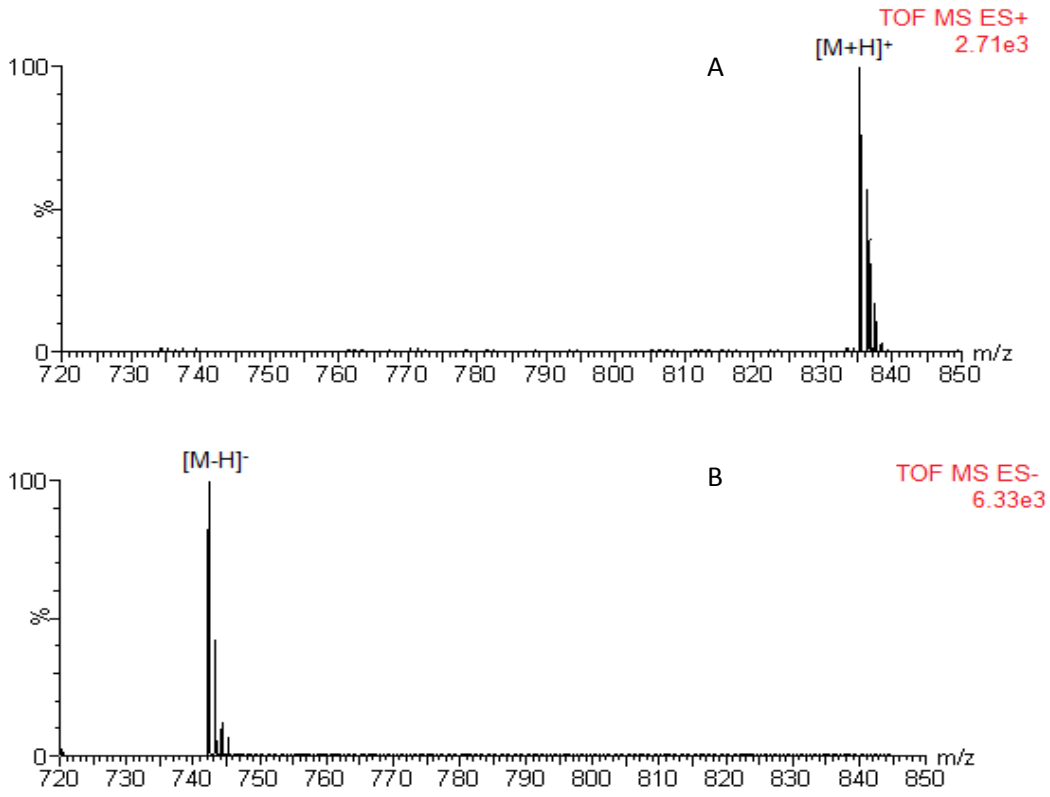


Figure 2.8: A) Positive ion-mode ESI-MS of peptide **4a** containing 4-pyridylmethyl ester protected glutamic acid. B) Negative ion-mode ESI-MS of peptide **5a**. Note, peptide **5a** was not observed by positive ion-mode ESI-MS using standard conditions and operating parameters *(163)*.

The work demonstrates the effectiveness of the 4-pyridylmethyl ester protecting group (abbreviated OMePyr) in switching the charge state of a peptide from negative to positive, thus making detection by positive ion mode ESI-MS possible.

Although peptides **5a** and **5b,** which lack any basic groups, were detected by negative ion mode, ESI-MS problems have been noted in the analysis of highly acidic sequences. In particular, the high net negative charge of these peptides causes them to attract and retain adventitious sodium ions. This leads to reduced sensitivity of the mass spectral analysis because the ion intensity is spread across a wide distribution of peptide–sodium complexes. As strong anion exchange

(SAX) is often the method of choice for purification of such sequences, sodium or other salts are necessarily introduced to the sample. Further, ion-pairing agents like TFA, which is widely used in peptide synthesis and purification, also interfere in negative mode ESI-MS as pre-ionized sodium/TFA adducts appear with very high intensities. In many cases, these ion pairs completely mask the presence of analyte in mass spectra.

Use of 4-pyridylmethyl ester side-chain protection in conjunction with the benzyl ester side-chain protection increases the hydrophobicity of otherwise hydrophilic peptides so that purification by RP HPLC is possible. Table 1 shows the retention times of three hexapeptides composed solely of glutamic acid synthesized using different protecting group strategies. Use of exclusively t-butyl ester protecting groups, which are removed during cleavage from the resin, yields a peptide with 11.2 min retention time using a standard gradient. Inclusion of three benzyl ester groups lengthens the retention time to 22.2 min whereas use of two benzyl esters in addition to one 4-pyridylmethyl ester gives a retention time of 19.3 min. Thus, the chromatographic properties of the peptide can be controlled with the judicious use of a combination of these two protecting groups.

Peptide	Retention time (min)
Ac-Glu-Glu-Glu-Glu-Glu-Glu-NH_2	11.2
Ac-Glu-Glu-Glu-Glu-Glu-Glu-NH_2 (OBzl) (OBzl) (OBzl)	22.2
Ac-Glu-Glu-Glu-Glu-Glu-Glu-NH_2 (OBzl) (OBzl) (OMePyr)	19.3

Table 2.1: HPLC retention times for hexapeptides composed of glutamic acid synthesized using different protecting group strategies. HPLC run conditions: Buffer A: 0.1% TFA (Trifluoroacetic acid) in acetonitrile, Buffer B: 0.1% TFA in water. Isocratic elution with 100% B for 2 min, gradient elution with 100% B to 0% B for 32 min, and Isocratic with 100% A for 5 min *(163)*.

Due to the "charge flipping" afforded by the 4-pyridylmethyl group, eluted peptides can be readily characterized using positive mode ESI-MS. After drying, hydrogenation yields the final peptide product without introduction of salts or excess TFA so characterization can be performed using negative mode ESI-MS.

Section 2.5 Conclusions

Problems associated with characterization can easily lead one to the conclusion that a synthesis has failed when in fact it has not. It is critical to develop a purification method that is compatible with ESI-MS. The method presented herein can be easily and quickly applied to the characterization of a variety of highly acidic peptides.

The 4-pyridylmethyl ester serves as a useful protecting group for glutamic acid and aspartic acid in solid-phase peptide synthesis and allows for the charge state of a peptide to be switched from negative or neutral to positive, thus allowing for characterization by positive mode ESI-MS. In conjunction with benzyl esters, the 4-pyridylmethyl ester allows for the chromatographic properties to be modified such that peptides that otherwise have very short retention times or no retention at all on a standard RP HPLC column can be retained. The OMePyr group is readily removed from peptides in high yield by hydrogenation. For larger sequences, combining the 4-pyridylmethyl ester protecting group with benzyl ester protection reduces the number of the former needed to produce a net positive charge and allows for purification by RP HPLC.

Chapter 3

Zn^{2+}-induced structural changes in Prot*α*(50-89)N50W studied using CD spectroscopy

Section 3.1 Introduction to Circular Dichroism Spectroscopy

Circular dichroism (CD) is the difference in the absorption of left-handed circularly polarized light (L-CPL) and right-handed circularly polarised light (R-CPL) and occurs when a molecule contains one or more chiral chromophores (light-absorbing groups):

$$\text{Circular dichroism} = \Delta A\ (\lambda) = A\ (\lambda)LCPL - A(\lambda)RCPL$$

where λ is the wavelength and A is the absorbance.

For CD measurements the most important features required for a sample are presence of chirality and chromophore group. The technique is extensively used to study chiral molecules of all types and sizes. CD can be used to study structural aspects of chiral molecules and their interactions with other molecules.

CD Instrumentation

Most commercial CD instruments are based on the modulation techniques introduced by Grosjean and Legrand. Light is linearly polarized and passed through a monochromator. The single wavelength light is then passed through a modulating device, usually a photoelastic modulator (PEM), which transforms the linear light to circular polarized light. The incident light on the sample switches between LCP and RCP light. As the incident light switches direction of polarization, the absorption changes and the difference in molar absorptivity can be calculated (Figure 3.1).

Figure 3.1: The instrumentation for a common CD spectrometer showing the polarization of light and the differential absorption of LCP and RCP light.

A primary use of CD is in analyzing the secondary structure or conformation of macromolecules, particularly proteins. As secondary structure is sensitive to its environment, temperature or pH, circular dichroism can be used to observe how secondary structure changes with environmental conditions or on interaction with other molecules. Structural, kinetic and thermodynamic information about macromolecules can be derived from circular dichroism spectroscopy.

The most widely studied circular dichroism signatures are the various secondary structural elements of proteins such as the α-helix, the β sheet, and the random coil.

Strengths and weakness of CD

The advantages of CD are derived from speed and convenience of the technique. Compared with the X-ray crystallography and NMR, CD can obtain a good quality of spectra both in near and far UV CD studies within 30 min time. Far UV CD studies, in particular, require only small amounts of material and since the technique is non-destructive, it is usually possible to recover most or all of the solution and hence one can conduct multiple experiments on the same sample. The ability to use cells of path lengths differing by several orders of magnitude means that it is relatively easy to study a very wide range of protein concentrations. This latter point represents an

important advantage of CD over Fourier Transform Infra Red (FTIR) spectroscopy which can give useful information about the secondary structure of proteins, but under a more restricted range of experimental conditions *(99)*. In addition, CD studies can be performed over a wide range of experimental conditions, including pH and temperature (with suitable attention to the requirement to avoid excess absorption or scattering of the incident radiation). Developments in instrumental design have now also made it possible to use stopped flow CD to study structural changes in proteins which occur on the timescale of a few tens of milliseconds *(100, 101)*.

CD of Proteins:

The most common use of CD in proteins is to compare global structural features of two macromolecules, or to analyze a single molecule under different conditions. This can be used simply to ascertain if a newly purified protein is correctly folded, determine if a mutant protein has folded correctly in comparison to the wild-type, or for the analysis of biopharmaceutical products to confirm that they are still in a correctly folded active conformation.

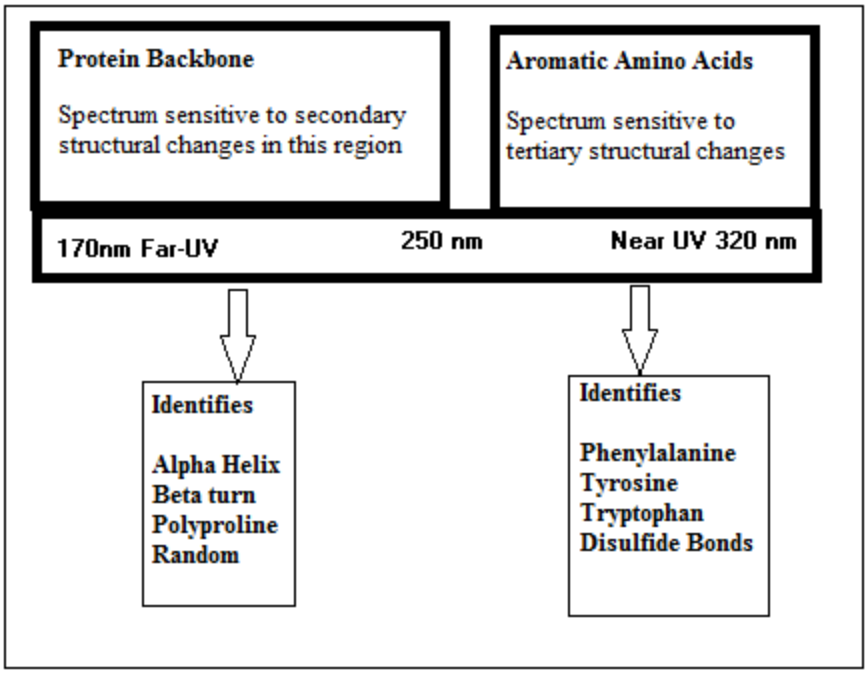


Figure 3.2: CD spectrometer showing the wavelengths and structural types it can identify in proteins.

Chromophores commonly observed in CD studies of proteins

CD can measure various aspects of protein structure. There are two types of optically active chromophores in proteins. In far UV CD the peptide bond is the principal absorbing group. Studies in this region can give information about the secondary structure. In near UV CD aromatic side chains of amino acids such tryptophan, tyrosine, and phenylalanine absorb in 250-290 nm range (Figure 3.2).

In addition, non-protein components or cofactors, such as pyridoxal-5'-phosphate, haem, flavin, bacteriochlorophyll, or added ligands may absorb in regions of the spectrum well separated from those of amino acids and peptide bonds. The CD signals in these regions can be used to provide

detailed information on the environment and possible interactions between these cofactors or ligands.

The contribution of the peptide bond and aromatic amino acid side chains in identifying the secondary structure of the protein will be discussed in detail.

The peptide bond

In the far UV region (typically 240 nm to 190 or 180 nm) the absorbing group is principally the peptide bond. There is a weak but broad $n \rightarrow \pi^*$ transition centered around 210 nm and an intense $\pi \rightarrow \pi^*$ transition at about 190 nm *(102)*. Studies of far UV CD can be used to assess quantitatively the overall secondary structure content of the protein (Figure 3.3).

Aromatic side chains

Based on the location of the side chains and type of amino acid containing side chain the secondary structure of a protein gives characteristic near UV CD spectrum. Each of the aromatic amino acids tends to possess a characteristic wavelength profile *(103)*. Tryptophan, a peak close to 290 nm with fine structure between 290 and 305 nm; tyrosine, a peak between 275 and 282 nm (the fine structure at longer wavelengths may be obscured by that from tryptophan); phenylalanine, sharp fine structure between 255 and 270 nm.

The types of information that can be obtained from CD of proteins and peptides include secondary structure content, conformational changes, and protein and peptide refolding.

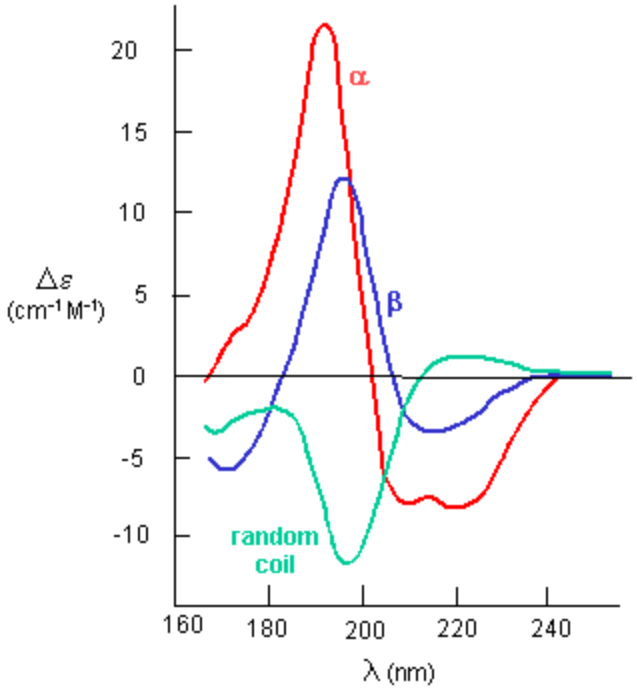


Figure 3.3: Far UV CD spectra of different secondary structures in proteins. The Figure shows the CD spectrum of α-helix, β-sheet, and random coil.

CD studies of intrinsically disordered proteins

Extended IDP's are disordered under physiological conditions because of the strong electrostatic repulsion and weak hydrophobic interactions. IDP's with their highly biased amino acid sequences might possess unusual conformational responses to changes in environments such as strong denaturants, temperature, pH, counter ions, osmolytes, binding partners, and molecular crowding *(104)*

Effect of Temperature

The analysis of the temperature effects on structural properties of several extended IDP's revealed that native coils and native pre-molten globules possess a so-called "turned out" response to heat. This is illustrated in Figure 3.4 where the temperature induced changes in the

far-UV CD spectrum of human α-synuclein is shown *(105, 106)*. The temperature dependence of the CD signal at 222 nm ($[\theta]_{222}$ characteristic wavelength of α-helix) for several extended IDPs (Figure 3.4, 3.5, 3.6) is shown. Figure 3.4 shows that at low temperature α-synuclein possessed a far-UV CD spectrum typical of an unfolded peptide chain. As the temperature was increased, the shape of the spectrum changed, reflecting the temperature induced formation of secondary structure. The spectrum also shows the presence of an isodichroic point (isosbestic point) at about 208 nm – that is a wavelength at which the total absorbance of the sample does not change during the heating process (or chemical process in the case of pH or binding, etc.). This indicates a two state process – presumably an unfolded state going to a more folded or collapsed state. This temperature inducing formation of α-helix is better illustrated by plotting $[\theta]_{222}$ versus temperature for α-synuclein (Figure 3.4 inset) *(105-107)*, caldesmon 636-771 fragment *(119)* (Figure 3.5), and phosphodiesterase γ-subunit (Figure 3.6) *(108) (109)*. The major structural changes observed over the range of 20 to 90 ^{o}C and further heating beyond 90 ^{o}C was accompanied by less pronounced effects. The structural heat induced changes in all these IDP's were completely reversible. Thus, an increase in temperature induces partial folding of IDP's, rather than the unfolding typical of ordered globular proteins, which was referred to as "turned out" response to heat.

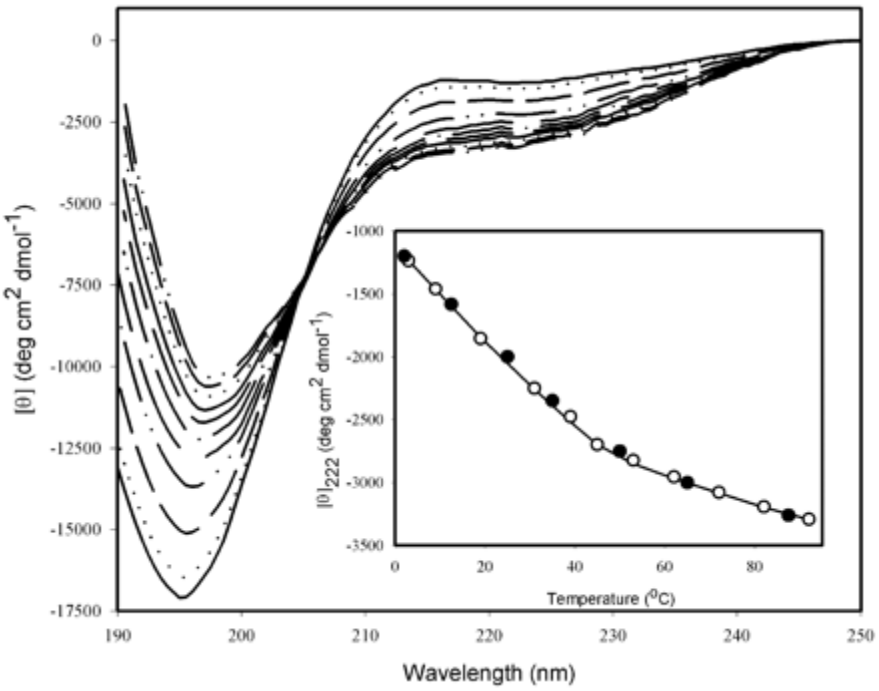


Figure 3.4: The temperature-induced changes in the far-UV CD spectrum of human α-synuclein. Far-UV CD spectra were measured at increasing temperatures: 3.0, 9.0, 19.0, 31.0, 39.0, 45.0, 53.0, 62.0, 72.0, 82.0, and 92.0 ^{o}C. The molar ellipticity of the peak at approximately 198 nm becomes more negative and shifts to lower wavelengths with increasing temperature *(105)*.

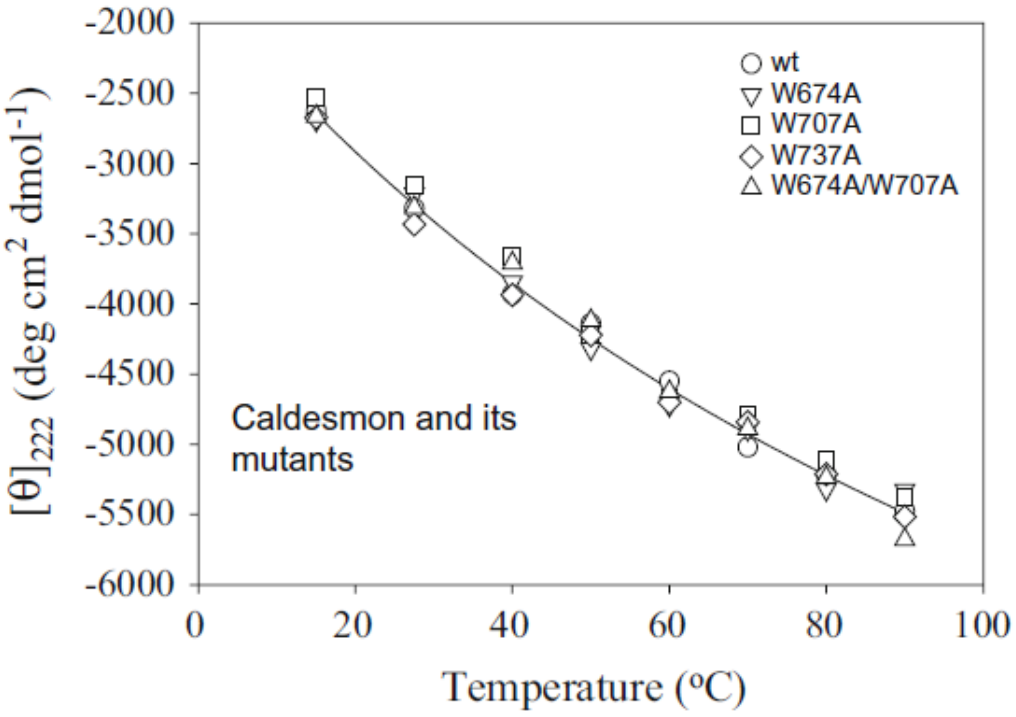


Figure 3.5: Effect of temperature on far UV CD spectra of chicken gizzard wild type caldesmon (636-771 fragment) and its mutants, depicted as $[\theta]_{222}$ vs. temperature *(110)*.

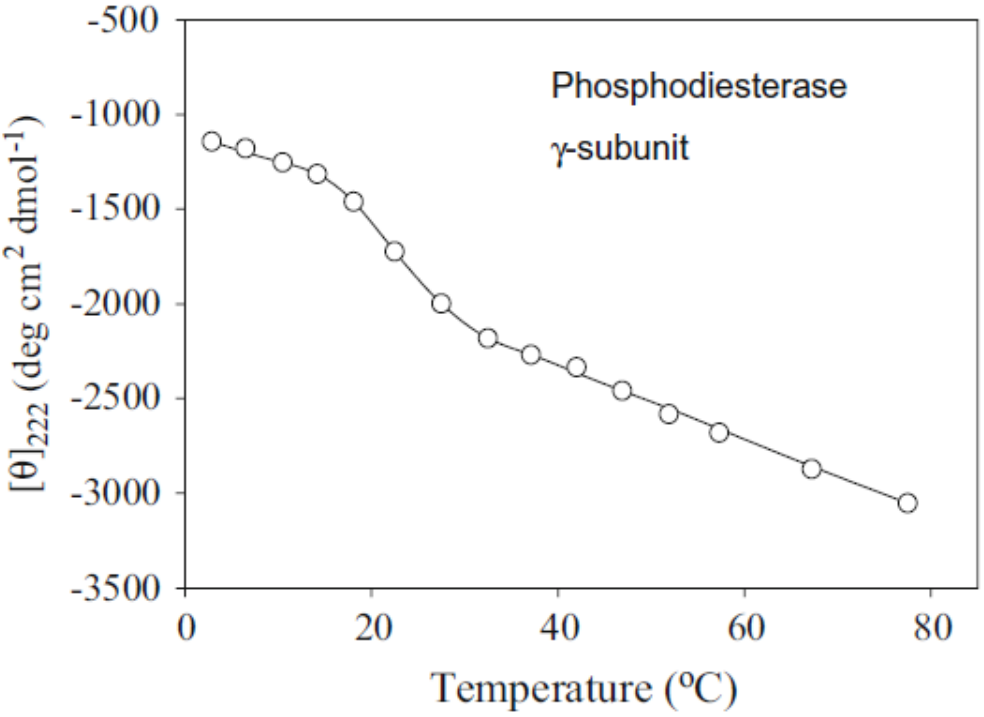


Figure 3.6: The temperature-dependence of $[\theta]_{222}$ of γ-subunit of phosphodiesterase *(108, 109)*.

Effect of pH

Extended IDP's are also characterized by the "turned out" response to changes in pH. Figure 3.8 represents the pH dependence of α-synuclein *(105, 106)* and prothymosin-α as monitored by CD *(59)*. It is clearly seen that there was a little or no change in the far UV CD spectra between pH 9.0 and 5.5. However, a decrease in pH from 5.5 to 3.0 resulted in a substantial increase in negative ellipticity at 220 nm. It has been established that the pH induced changes in far UV CD spectra of these two proteins were completely reversible and consistent with the formation of partially folded pre-molten globule like conformation. Similar pH-induced structural transformations have been described for such extended IDPs as pig calpastatin domain I *(111)*, histidine rich protein II *(112)*, and naturally occurring human peptide LL-37 *(113)*. These observations show that a decrease in pH induces partial folding of extended IDPs due to

minimization of their large net charge present at neutral pH, thereby decreasing charge/charge intramolecular repulsion and permitting hydrophobic-driven collapse to the partially folded conformation. In contrast to this, for a folded protein a pH drop to acidic range (pH 1-3) would result in protein denaturation and loss of tertiary structure. The net charge on the protein due to the titration of all the ionizing groups leads to intramolecular charge-charge repulsion, which is sufficient to overcome the attractive forces (mostly hydrophobic and dispersive) resulting in at least partial unfolding of the protein.

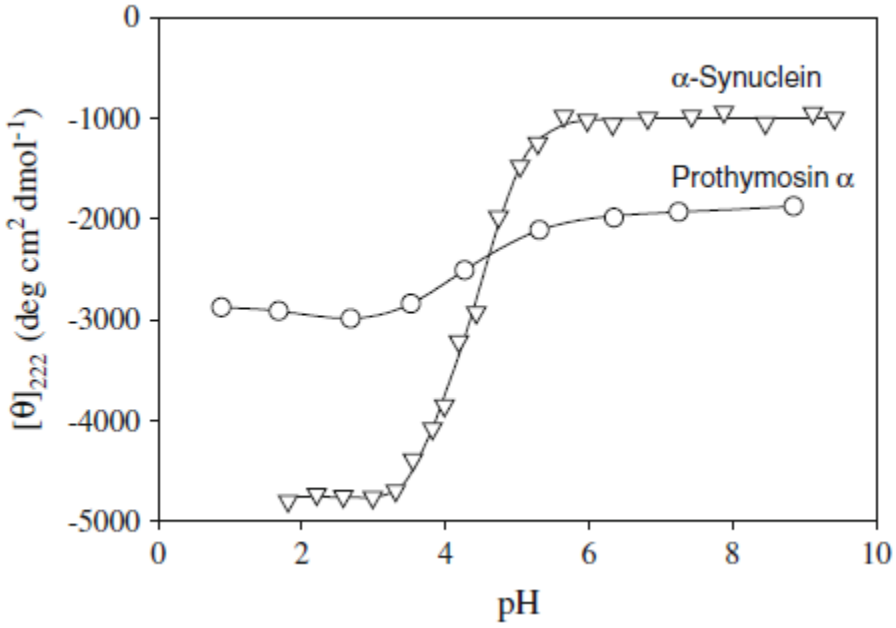


Figure 3.7: The effect of pH on structural properties of IDPs. The pH dependence of $[\theta]_{222}$ for α-synuclein and Protα *(30, 59, 105)*.

Effect of counter ions

Interactions of IDPs with ligands that affect their net charge and net hydrophobicity often allow the protein to adopt more regular structure approximating the structure seen in more ordered proteins. This hypothesis is illustrated by seven IDPs in the presence and absence of cationic species. In Figure 3.9 the black circles correspond to the ligand-free proteins whereas open

circles represent proteins complexed with particular metal cations. The solid line represents the border between IDPs and folded proteins *(16, 43)*. By this analysis metal binding can induce structure and this idea is well supported by Figure 3.10. Several cations (monovalent, divalent, and trivalent) were shown to induce conformational changes in α-synuclein and transformed this extended IDP into partially folded conformation. The folding strength of cations increases with increase in ionic charge density. Human anti-bacterial protein LL-37, a natively unfolded protein with a high net positive charge was shown to be folded in the presence of several anions *(113)*.

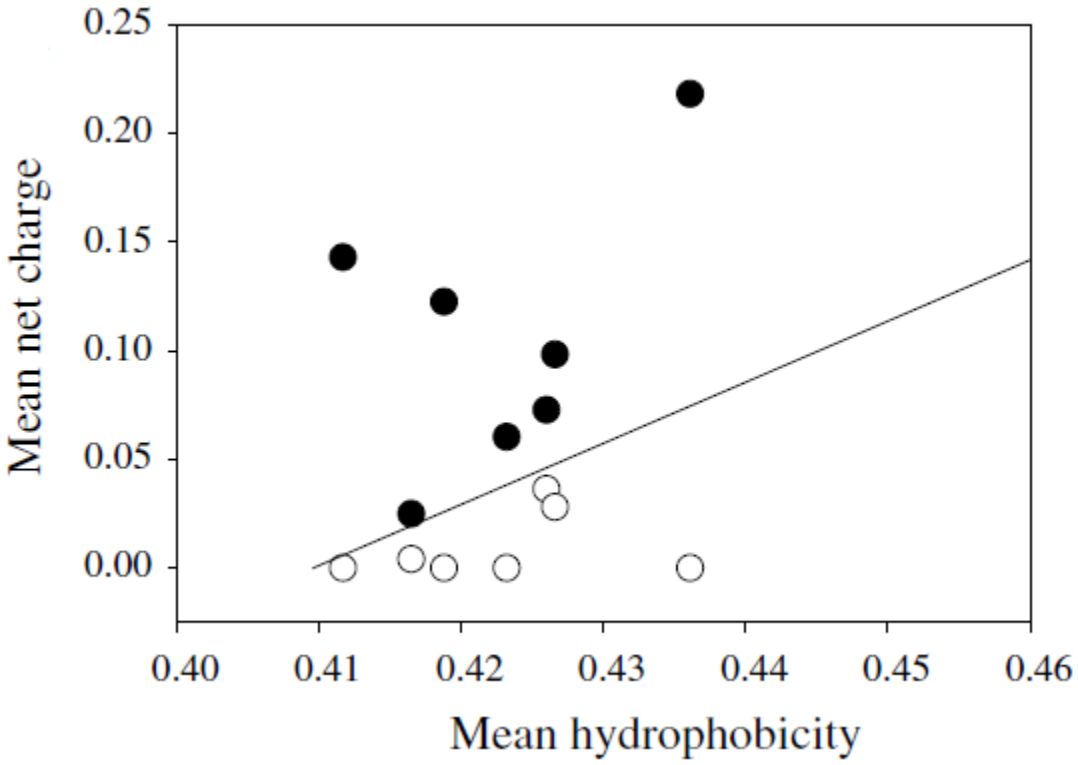


Figure 3.8: The effect of counter ions on the ratio of mean net charge to mean hydrophobicity of IDPs. Charge-hydropathy plot for seven extended IDPs, ostecalcin, osteonectin, α-casein, HPV16E7 protein, calsequestrin, manganese stabilizing protein and HIV-1 integrase in their apo-form (black circles) and metal-saturated (white circles) forms *(16)*. Ca^{2+}, Mn^{2+}, and Zn^{2+} are the cations used in the study.

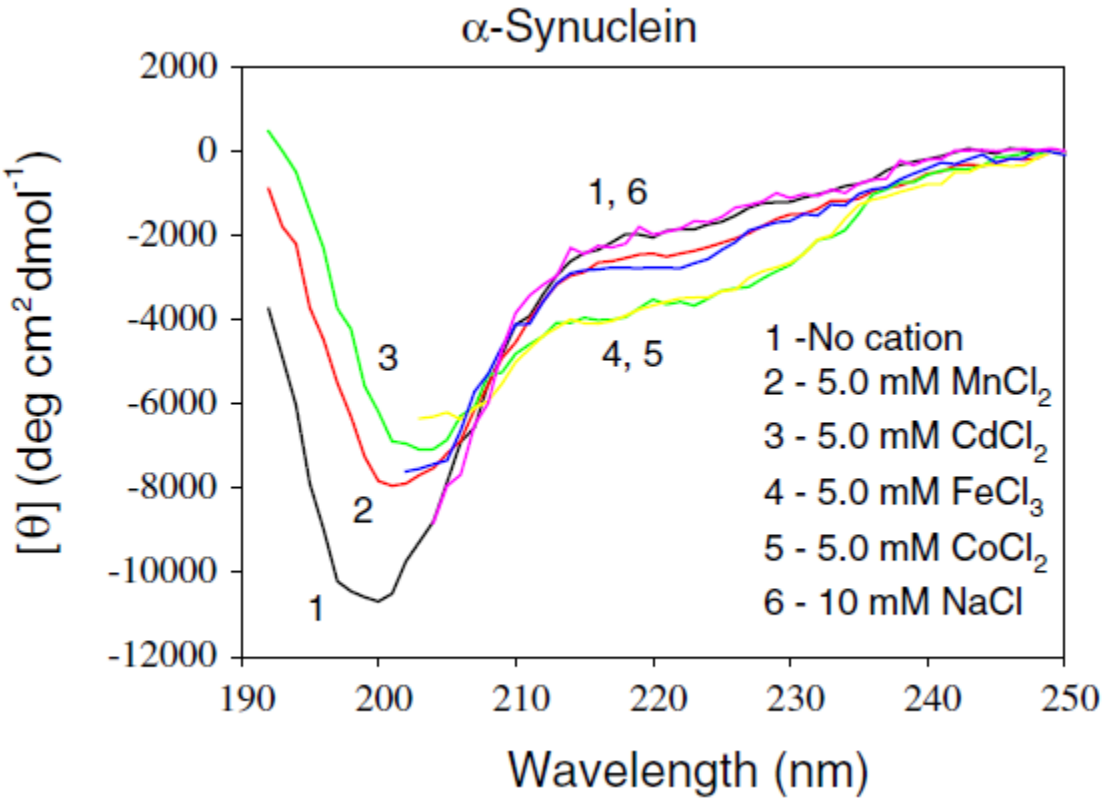


Figure 3.9: The far-UV CD spectra of α-synuclein in the presence of various metal cations *(105)*.

Elucidation of the molecular features and biophysical properties of IDPs with environmental changes will provide more insight into their function and potentially novel mechanisms of action, thus contributing to a more complete understanding of IDP's metal binding. It will be interesting to see if IDPs as a whole make use of fundamentally different metal binding motifs than their structured counter parts. Information on even a few proteins of this class will give insight into whether a lack of structure is a way to confer multi-functionality to a single protein.

Section 3.2 Research Strategy

Protα(50-89)N50W is used as a model peptide to carry out CD studies. Though 48-110 region of the protein is found to be the metal binding region, like mentioned in section 1.5 most of the Zn^{2+} binding ligands (long stretches of Glu and Asp) are located most likely in the central 50-89 region of the protein. In general, the most common ligands for Zn^{2+} are cysteine and histidine. However, Protα completely lacks any of these residues. The next possible ligands are carboxylate groups in the side chains of glutamic and aspartic acid *(114, 115)*.

So the 50-89 segment of Protα was chosen as a model peptide for the CD studies. Below is the sequence of Protα(48-100) and Protα(50-89).

A.) A D N E V D E E E E E G G E E E E E E E E G D G E E E D G D E D E E A E S A T G K R A A E D D E D D D V D T K K Q K T D E D D

B.) N E V D E E E E E G G E E E E E E E E G D G E E E D G D E D E E A E S A T G K R

Sequence 3.1: Sequence of Prothymosin showing A. 48-100 segment with 50-89 region highlighted. B. The central acidic region from residues 50-89.

Specific Aim 1.2 Determine the structural changes in the central segment of Protα in the presence of environmental factors like Zn^{2+} and temperature

The Zn^{2+} study using Protα(50-89)N50W will provide insight into the amount of structural change that occurs in this region versus in the full peptide. IDPs are known to gain structure at increased temperatures *(116)* and it will be interesting to see if this is the case with Protα. This study would show if the character of the structure adopted by the peptide is similar to that of the full length Protα when it binds Zn^{2+}.

1.2 A. Zn^{2+} induced structural changes in Prothymosin-α (50-89 N50W) observed by CD Spectroscopy

We hypothesize that the structural changes in the protein occur primarily in the central region. Our first aim was to look for the structural changes in the central segment of the protein in the presence of Zn^{2+} by using CD. The central 50-89 segment was used to carry out these studies. CD measurements were made by titrating zinc into the peptide solution. Spectra obtained from these experiments were compared to the CD spectrum of the full length protein *(43)*

In this direction, our initial step was to determine the optimum concentration of the peptide that would give us a good signal on CD. The peptide at the determined concentration was then titrated with varying amounts of zinc. CD spectra were obtained for each addition of zinc. The data obtained was overlaid to look for the structural changes in the peptide at different concentrations.

Experiments:

All CD measurements were carried out on a Jasco J-815 CD Spectrometer using a 0.1 mm cylindrical quartz cell, 250 nm to 190 nm wavelength at 25 ^{o}C. The peptide and zinc chloride solutions were prepared in a 50 mM phosphate buffer and adjusted the pH to 7.4.

Expected results:

The peptide and the full Protα are highly negatively charged and have strong electrostatic repulsion resulting in mostly unfolded structure to the protein. Cationic metal ions like Zn^{2+} are expected to establish electrostatic interactions with acidic side chains resulting in reduction of repulsion and thereby causing folding of the peptide. Since the full protein has been shown to

adopt a more ordered structure upon addition of Zn, we expect zinc to induce same transition in the peptide also.

1.2 B. Temperature-dependent structural changes in Prothymosin-α(50-89)N50W

A large number of CD studies have demonstrated that a structural change takes place in IDPs with increasing temperature which most likely reflects formation of transient α-helices *(116)*. The studies show that the temperature-induced structural change is common among IDPs and is accompanied by a contraction of the conformational ensemble. Structural changes in Protα as a function of pH and Zn^{2+} have been studied before. However temperature dependent CD studies of Protα have not been reported.

Experiments

Initially optimum concentration of Protα(50-89)N50W that gives us good signal was determined to be 0.0168 mM. A cell with 1 mm pathlength was used. The peptide was then heated from 20-80 ^{o}C. Once the peptide reached 80 ^{o}C it was cooled down to 20 ^{o}C. The heating and cooling processes were carried out by a 10 ^{o}C increase and decrease respectively for each run. The system was equilibrated for 15 min for every 10 ^{o}C raise in temperature.

In the second set of runs the peptide with 50 eq of Zn^{2+} added to it was heated and cooled again. IDPs are shown to attain secondary structure at high temperatures. Since that was true in the case of Protα(50-89)N50W, structural changes of Protα(50-89)N50W under the influence of two factors; Zn^{2+} and heat were observed. These experiments help us understand the effect of temperature in the presence of metal cations.

1.2 C. Zn^{2+} induced structural changes in polyglutamic acid sequence

Polyglutamic acid (polyE) sequences are known to bind Zn^{2+} *(43)*. The effects of Zn^{2+} on the structural features of a polyglutamic acid containing the same number of residues as Protα(50-

89)N50W were studied using far UV CD. This helps us answer if the structural changes in Protα and its peptides occur solely due to electrostatic interactions with Zn^{2+} or if there is sequence specificity that causes the structural change.

Ac-WEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEKR-NH_2

Sequence 3.2: Polyglutamic acid sequence having same number of residues as Protα(50-89)N50W.

A polyE sequences (shown above sequence 3.2) has been synthesized. Two spectra of polyE one with 200 eq of Zn^{2+} and the other without Zn^{2+} were obtained.

Expected Results:

We expect to see no change in the secondary structure of polyE in the presence of Zn^{2+} because of sequence specificity of Zn^{2+} to Protα which is not present in PolyE.

Aggregation

Aggregation or self-association is a characteristic property of intrinsically disordered proteins. As Protα belongs to the class of "natively unfolded" proteins, increase in temperature may accompany self-aggregation. To test this data obtained at same temperature in the heating and cooling cycle will be overlaid. If there is no self-aggregation in the sample both spectra overlay upon each other.

Section 3.3 CD Measurements

Metal and temperature-induced structural changes in the central metal binding region (50-89) N50W of prothymosin-α have been investigated. The following CD studies were carried out.

1.2 A. Zn^{2+} induced structural changes in prothymosin-α(50-89)N50W observed by CD Spectroscopy

The far-UV CD spectrum of prothymosin-α (50-89)N50W in 50 mM phosphate buffer was recorded at pH 7.5 and at room temperature. The pH and temperature were kept consistent with CD runs of the full length protein collected by Uversky *et al.* so the results can be directly compared *(43).* The concentration of the peptide used was 0.5 mM and the spectra were collected in the wavelength range of 190-260 nm. The peptide exhibits a maximum in the CD spectrum at about 220 nm and a minimum at 198 nm which is consistent with random coil conformers. With the addition of Zn^{2+} a significant gain in negative ellipticity around 200 nm and a negative shoulder around 222 nm were observed suggesting Zn^{2+} induced formation of secondary structure in the peptide. The negative shoulder at 222 nm is a characteristic wavelength of α-helix fomration in proteins. Figure 3.10A represents CD spectra of Protα(50-89)N50W peptide titrated with Zn^{2+}. Figure 3.10B represents the corresponding photomultiplier tube (PMT) high voltage values HT [V]. As recommended by Jasco, 400 V was selected as the threshold beyond which data may be considered excessively noisy and thus unreliable. This occurs at approximately 198 nm for all the CD spectra collected in this study. It should be noted that there is an isodichroic point at approximately 206 nm indicating a two-state transition. The peptide reached saturation at approximately 75 eq of Zn^{2+} which is same as observed for the full protein

(43). This supports the conclusions that majority of the metal added binds to the central 50-89 region.

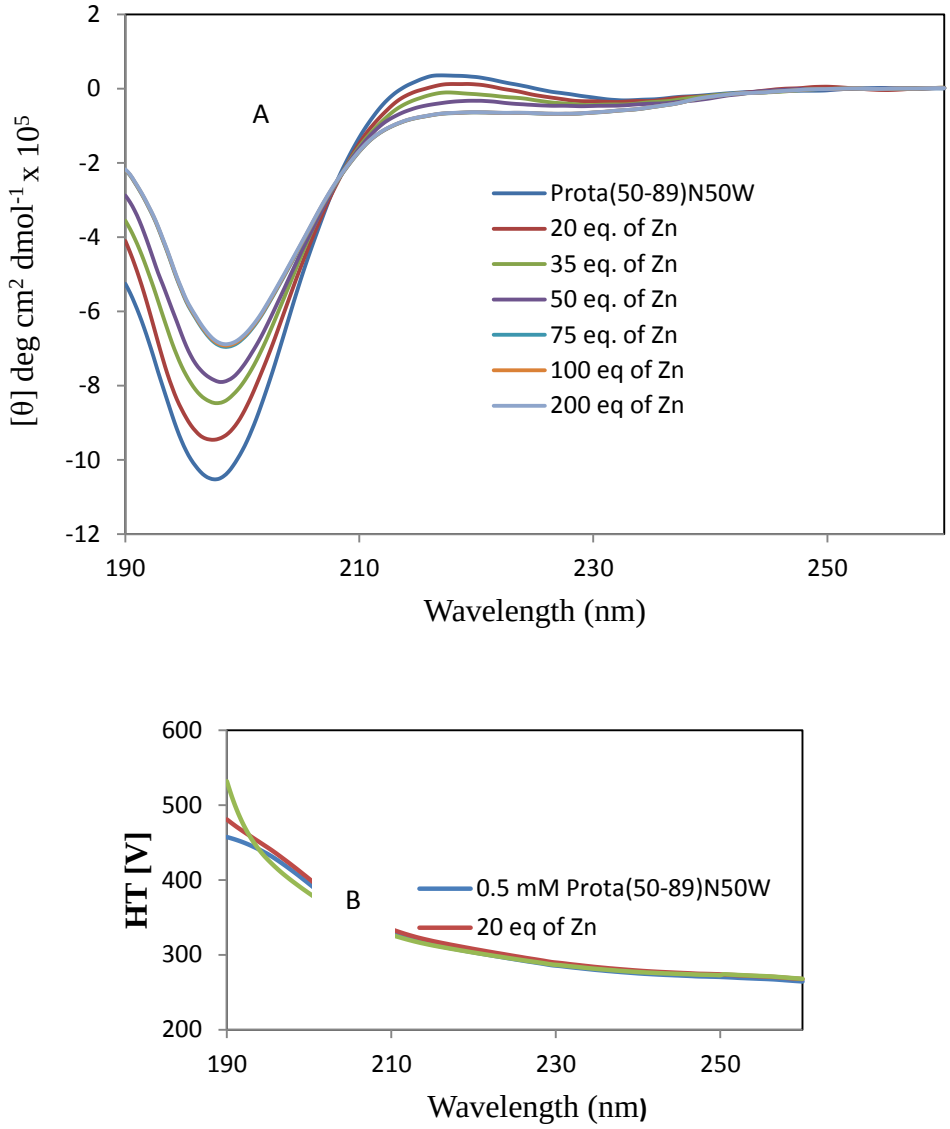


Figure 3.10: A. Zn^{2+} induced structural changes in Protα(50-89)N50W far-UV CD spectrum. **B.** Corresponding HT[V] shown for peptide with minimum and maximum concentrations of Zn^{2+}

(the HT[V] of spectra recorded with all other concentrations of Zn^{2+} lie within the minimum and maximum HT[V] range shown here).

1.2 B. Temperature-dependent structural changes in prothymosin-α(50-89)N50W

Initially, the far-UV CD spectra was recorded for Protα(50-89)N50W in 50 mM phosphate buffer at physiological pH and room temperature (25 ^{o}C) to assess its structure. The CD spectrum of the peptide had a maximum at 220 nm and a minimum at 198 nm which indicated a random coil conformation. The temperature dependence of the secondary structure of the protein was investigated from 20 to 80 ^{o}C. The data shows that with increasing temperature there is a gain in negative ellipticity around 222 nm and loss at around 200 nm which is consistent with the assessment that the peptide is attaining structure when heated. A negative ellipticity at 222 nm is a characteristic feature seen for α-helixes of proteins. This data suggests a heat-induced folding of the Protα peptide chain into an α-helical structure.

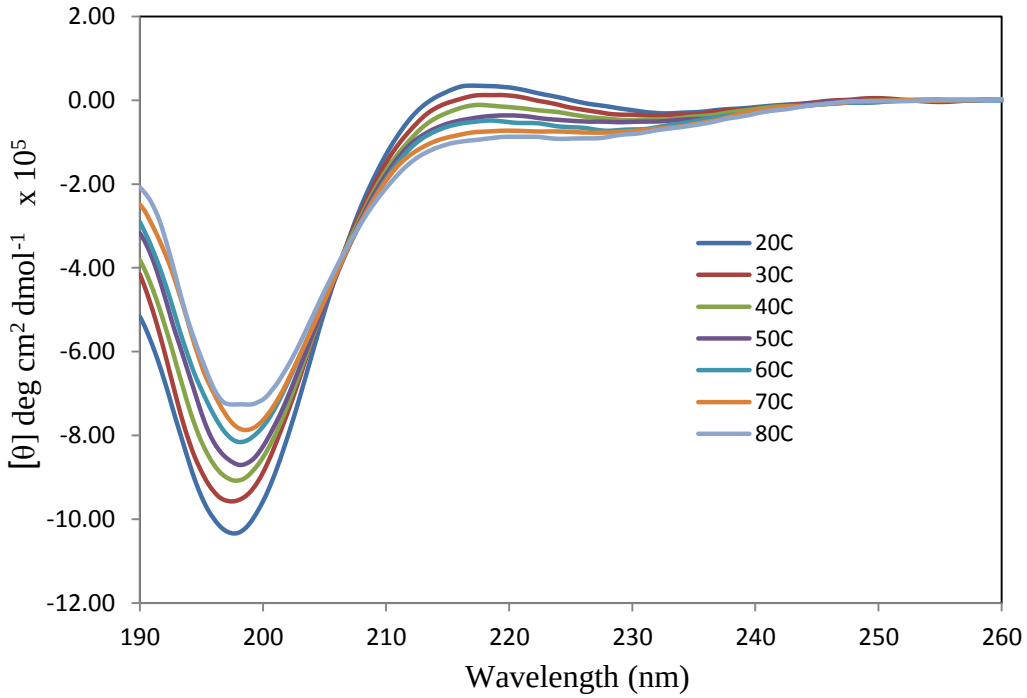


Figure 3.11: Far-UV CD spectrum of Protα(50-89)N50W collected in 10 ^{o}C increments. The spectrum shows the peptide adopts structure with increasing temperature.

Subsequently, temperature dependence of Protα-Zn complex was observed. 50 eq of Zn^{2+} was added to the peptide sample and the peptide-metal sample was heated from 20 to 80 ^{o}C at 10 ^{o}C increments. Figure 3.12 shows similar changes in ellipticity of the spectrum as observed in Figure 3.11. There is a gain in ellipticity at 222 nm and loss at 200 nm with increase in temperature. This data suggests that the peptide attains structure in the presence of Zn^{2+} when heated.

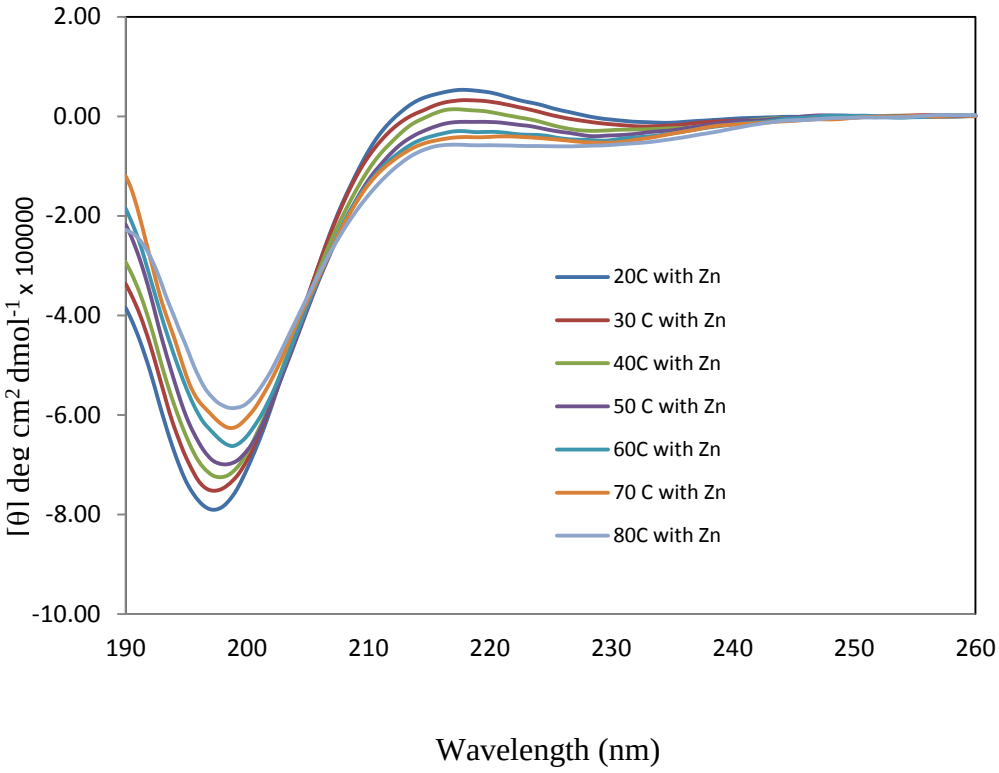


Figure 3.12: Far-UV CD spectrum of Protα(50-89)N50W with 50 eq of Zn^{2+} heated from 20-80 ^{o}C in 10 ^{o}C increments. The spectrum shows the peptide adopts structure in the presence of 50 eq of Zn^{2+} and with increasing temperature.

An overlay of spectra collected at different temperatures (20-80 ^{o}C) with peptide (only heat) and spectrum collected at respective temperatures with peptide and 50 eq of Zn^{2+} (heat and metal) showed that the peptide has extremely similar structures in both the cases. However, the amount of folding is less in Zn^{2+} present sample.

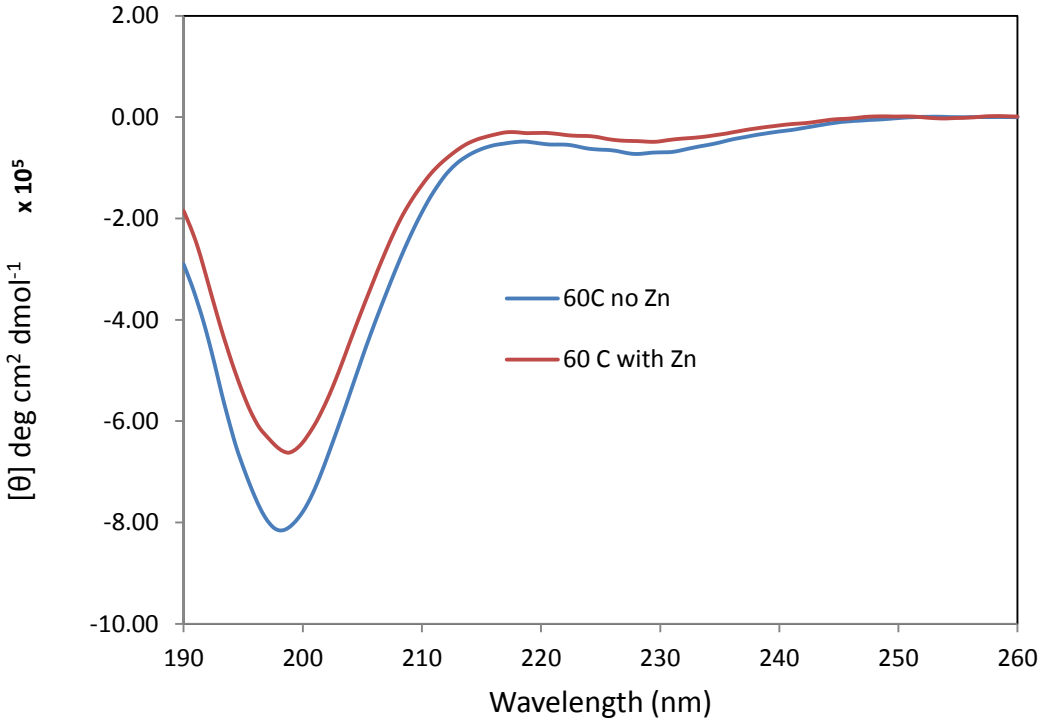


Figure 3.13: Far UV CD spectra of Protα(50-89)N50W at 60 ^{o}C with and without Zn^{2+}. The spectral overlay shows that the peptide adopts an overall similar structure in both the cases.

1.2 C. Zn^{2+} induced structural changes in the polyglutamic acid sequence

The structural changes in acidic peptides in the presence of Zn^{2+} have been thought to solely occur due to electrostatic interactions between Zn^{2+} and negatively charged residues in the peptides. We hypothesize that the ionic radius, high charge density, and sequence specificity of Zn^{2+} also plays an important role in the specificity of metal binding in the protein. To test the

hypothesis we synthesized a polyglutamic acid sequence of same length as Protα(50-89)N50W peptide (Sequence 3.1). The polyglutamic acid was designed in a way that residues that are found important in Protα(50-89)N50W were retained. The tryptophan at the N-terminal end of the peptide helps to measure concentration on UV. The KR residues at the C-terminus are found to be important for Protα(50-89)N50W activity. It was reported that Zn^{2+} binds specifically to polyglutamic acid (42). However, the study does not say if the binding induces a structural change.

The CD spectrum of polyE sequence exhibited a maximum at 220 nm and minimum at 198 nm which is consistent with random coil conformation. With the addition of large molar excess of (200 eq) Zn^{2+} to the peptide, there was no significant structural change observed. These results support the idea that high charge of Protα(50-89)N50W is not the sole factor responsible for the folding of the peptide in the presence of Zn^{2+}. The conformational change induced by Zn^{2+} binding is sequence specific.

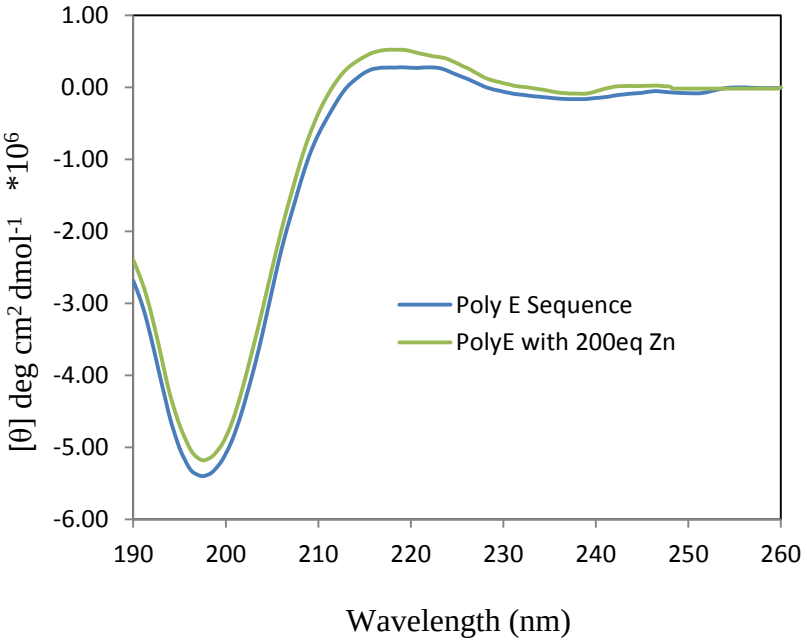


Figure 3.14: Far UV CD spectrum of polyglutamic acid in the presence and absence of Zn^{2+}. The peptide was treated with large molar excess of Zn^{2+} (200 eq.).

Over 50% of Protα(50-89)N50W comprises Asp and Glu and differs significantly from the polyglutamic acid sequence in terms of charge. The polyglutamic acid sequence has a greater overall charge than Protα(50-89)N50W which might inhibit folding based on increased electrostatic repulsion. A scrambled sequence of Protα(50-89)N50W (sequence 3.3) with same amount of overall negative charge as the peptide was synthesized and the structural characteristics of the sequence were studied in the presence of Zn^{2+}.

Ac-EEDEEEEKDEEAGDEEEDEERGGEEAGEVGEEEDESETG-NH_2

Sequence 3.3: Scrambled Protα(50-89)N50W sequence. The net charge of sequence remains the same as the peptide.

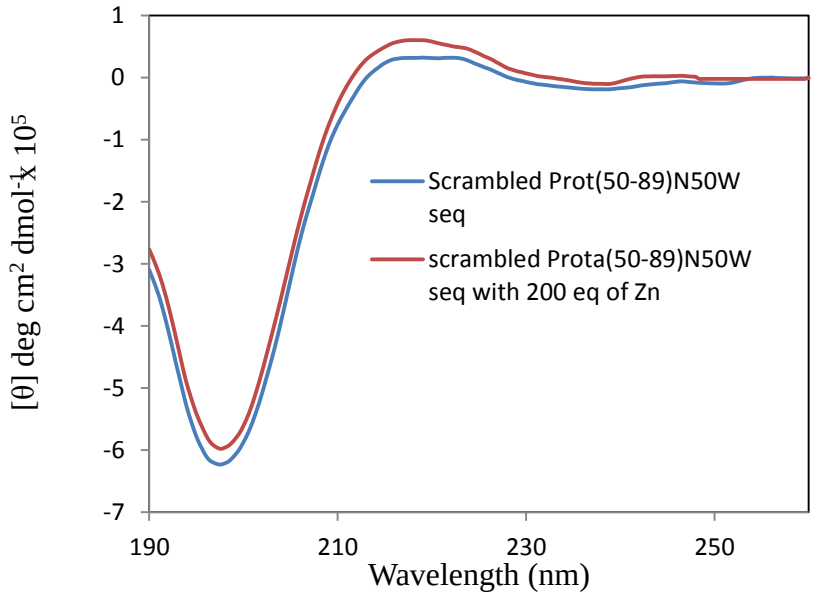


Figure 3.15: Far UV CD spectrum of scrambled sequence in the presence and absence of Zn^{2+}. The peptide was treated with large molar excess of Zn^{2+} (200 eq.).

Aggregation or self-association is a characteristic property of denatured proteins *(24–29),* and it has been shown that in some cases association induces more structure and stability in partially-folded intermediates *(28)(29).* To be sure there was no aggregation in any of the CD runs the far UV CD spectra of Protα(50-89)N50W at 25 ^{o}C, 80 ^{o}C and back at 25 ^{o}C were collected (Figure 3.16). The two spectra obtained at 25 ^{o}C before and after heating the sample overlay on each other indicating that the heat induced structural changes in the peptide are reversible and no aggregation occurred.

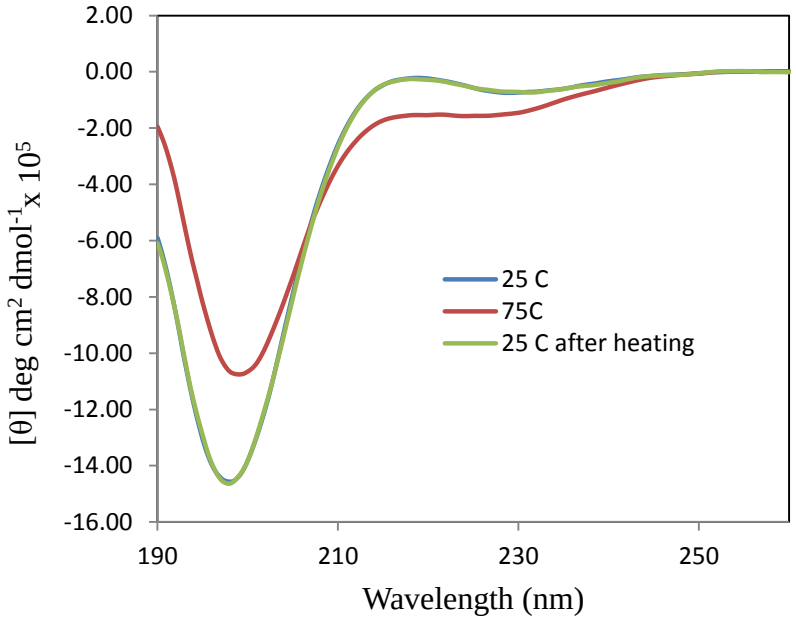


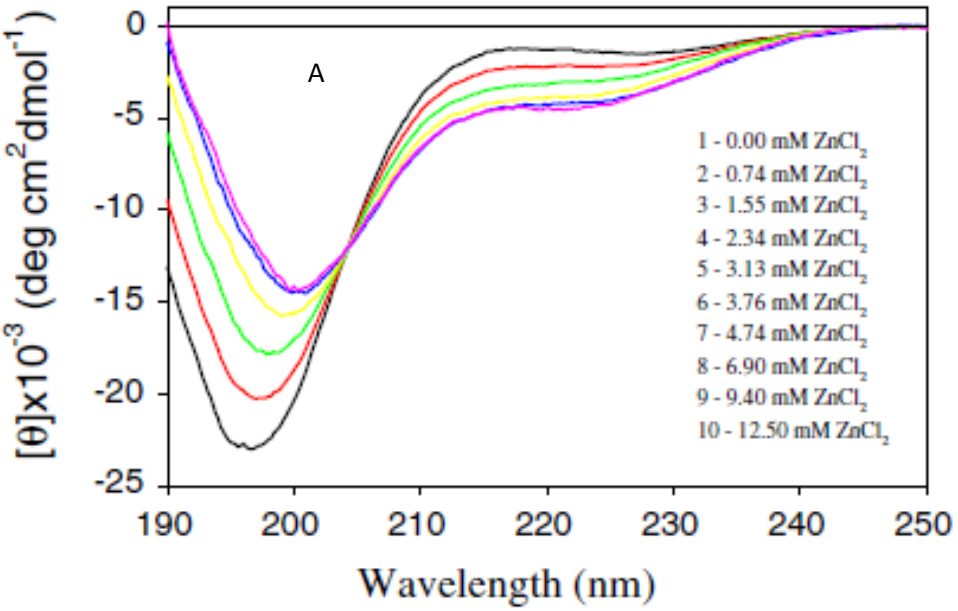
Figure 3.16: Far UV CD spectra showing that the secondary structural changes formed on heating and cooling of the peptide are reversible. There was no aggregation of the peptide observed. Note, that the two spectra collected at 25 ^{o}C are identical so the blue trace is overlaid by the green trace.

Discussion

The spatial structure of a protein molecule is often maintained by interaction with specific ligands *(117)*. This is true for Protα also. The data shown in literature tells us that zinc is a

specific ligand for human Protα and causes adoption of secondary structure in the protein upon binding *(48).* The interaction of Protα and Zn is said to be specific, since of the several divalent cations (including Ca^{2+}, Mg^{2+}, Mn^{2+}, Cu^{2+}, and Ni^{2+}) examined only Zn^{2+} showed the ability to induce substantial structural rearrangement in Protα.

Moving a step forward, an NMR study on Zn^{2+} binding of Protα has revealed that the 48-110 region of the protein is the Zn^{2+} binding region. There is still a need to explore the Zn^{2+} binding region of the protein in a more detailed manner to answer several questions left unanswered regarding metal binding in IDPs.



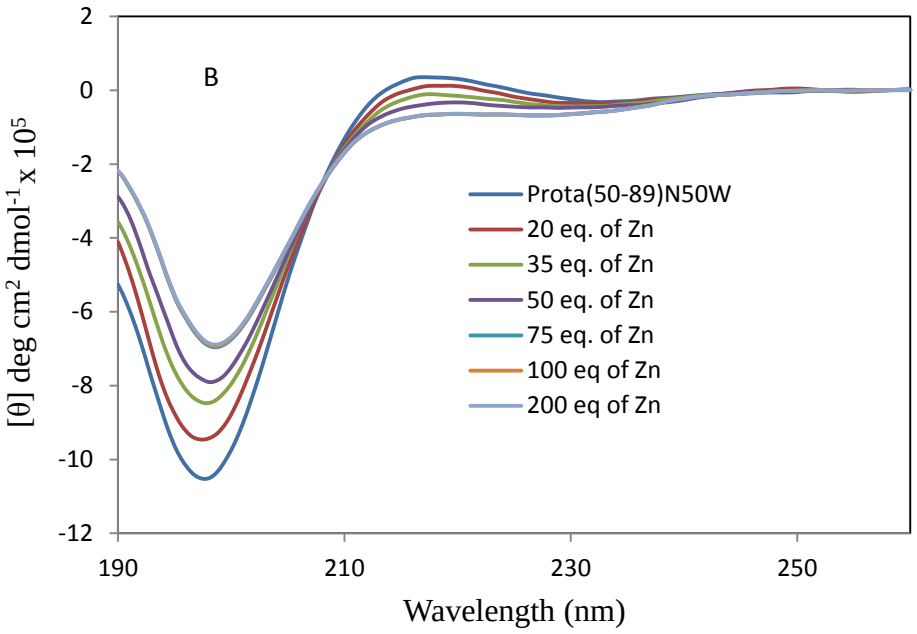


Figure 3.17: Far UV CD spectra of Protα showing Zn^{2+} titrations in A) 0.05 mM of full length Protα showing a change in the secondary structure with increasing concentration of Zn^{2+} and in B) Protα(50-89)N50W showing the similar structural change as the full protein in the presence of Zn^{2+} *(43, 118)*.

Figure 3.17 shows that full length Protα and the Protα(50-89)N50W peptide had a random coil like conformation at physiological pH and room temperature. When Zn^{2+} was titrated into them the spectra for both the samples showed an increasing negative ellipticity at 222 nm which is a characteristic of α-helical formation. This suggests that the shape adopted by the full Protα and Protα(50-89)N50W as a function of Zn^{2+} are very similar. Thus the majority of the structural change appears to be occurring in the central region due to interaction with Zn^{2+}. It was also found that the peptide like the full protein reached saturation at 75 eq of Zn^{2+} indicating that majority of the Zn^{2+} added to the full protein binds to the central acidic region (50-89). This

implies that the metal binding studies carried out using the 50-89 segment of Protα should model the full length protein and that the peptide could be used for Zn^{2+} studies instead of the full protein.

A second environmental factor that was studied here was temperature. A characteristic temperature-dependent spectroscopic change has been observed in Protα(50-89)N50W. Studies indicate that the peptide undergoes a structural change (similar to what was seen with Zn^{2+} binding) when heated from 20-80 ^{o}C both in the presence and absence of Zn^{2+}. The temperature induced changes of Protα(50-89)N50W were reversible with no detectable association of the protein. In other words, we can assume that the observed changes are of an intra- rather than an intermolecular character. The effects of elevated temperature may be attributed to increased energy to overcome repulsive electrostatic interactions and thereby the peptide experiences a collapse indicating folding.

Both Zn^{2+} and temperature cause Protα(50-89)N50W peptide to adopt structure. Further, the very close resemblance of the CD spectra obtained as a function of Zn^{2+} and temperature suggest the overall structures in both the cases may be similar.

To further investigate the nature of Zn^{2+} induced folding the following question was asked: Does Zn^{2+} induced folding of Protα occurs due to electrostatic interaction between the protein and the metal? To answer this a polyglutamic acid sequence was synthesized and the Zn^{2+} binding of the sequence was tested in the presence of large molar excess of Zn^{2+} (200 eq. of Zn^{2+}) (sequence 3.4) . The sequence showed no change in the structure upon addition of metal (Figure 3.14). This suggests that having a region with a high net negative charge is not sufficient for Zn^{2+}-induced structural changes. Thus there is some sequence specificity needed for Zn^{2+}-induced structural changes to occur in an acidic peptide.

The net negative charge of Protα(50-89)N50W is less compared to the polyglutamic acid sequence used for our study (sequence 3.5).

Ac-WEVDEEEEEGGEEEEEEEEGDGEEEDGDEDEEAESATGKR-NH_2

Ac-WEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEKR-NH_2

Sequence 3.4: Protα(50-89)N50W (on the top) and polyglutamic acid sequence (on the bottom). The negative residues are highlighted in red.

To match up the overall charge of the sequence with Protα(50-89)N50W CD spectra of a scrambled sequence of Protα(50-89) has been obtained in the presence and absence of Zn^{2+} sequence (sequence 3.4). The spectra showed that no major structural changes indicating Zn^{2+} induced structural changes in Protα(50-89)N50W are sequence specific (Figure 3.15).

Highly negatively charged proteins like Protα are unfolded in large part due to the electrostatic repulsion between the negatively charged residues. When an oppositely charged metal cation is introduced to the protein the charged side chains on the protein are neutralized and can transfer charge to the metal cation, resulting in neutralization of the positive charge on the cation (ligand to metal charge transfer, LMCT) *(114, 115)*. In the case of Protα the same happens and resulting in binding of the metal cations.

However, it may be asked how is it that of all the metals added (Mg^{2+}, Zn^{2+}, Ca^{2+}) and that can associate with Protα, only Zn can induce a structural change. So the next question of interest is why does the peptide show a structural change selectively with Zn^{2+}? Although Mg^{2+}, Zn^{2+}, Ca^{2+} are all divalent metal cations and ionic radius of all the three cations fall in the same range the

structural change in the peptide occurs only with Zn. Factors that govern selectivity of a metal are charge (valence state), ionic radius, and charge accepting ability. All the three metal cations (Mg^{2+}, Zn^{2+}, Ca^{2+}) have same charge and similar ionic radius (Mg^{2+} 0.72 Å, Zn^{2+} 0.74 Å, Ca^{2+} 1.0 Å). So the charge accepting ability of the metal is the only factor in this case that governs the selectivity of Zn^{2+} over other metal cations in this case. Hybridization of 3d and 4s orbitals of Zn^{2+} increases the charge acceptance ability of Zn^{2+} from the ligand, which is a carboxylate in this case. This is due to the fact that the energy of the ligand's highest occupied molecular orbital (MO) is close to that of the lowest unoccupied MO of the metal, thus leading to charge neutralization and folding of the protein in the presence of Zn^{2+}. In contrast, for other divalent metal cations like Mg, and Ca hybridization occurs between s and p orbitals and cannot accept as much ligand charge as the hybridized 3d 4s Zn^{2+} orbitals due to the greater energy difference *(115, 162).*

The stability of the Zn complex compared to the Mg and Ca complexes with the same ligand is explained by the binding gas-phase free energies (ΔG) for IIA and IIB metals ($M^{q+} + nH_2O \rightarrow [M(H_2O)_n]^{q+}$). Zn^{2+} complex has a ΔG of -95.4 kcal/mol which is much higher than the binding free energies of Mg^{2+} and Ca^{2+}, thus explaining the higher stability of Zn^{2+} complexes *(162).*

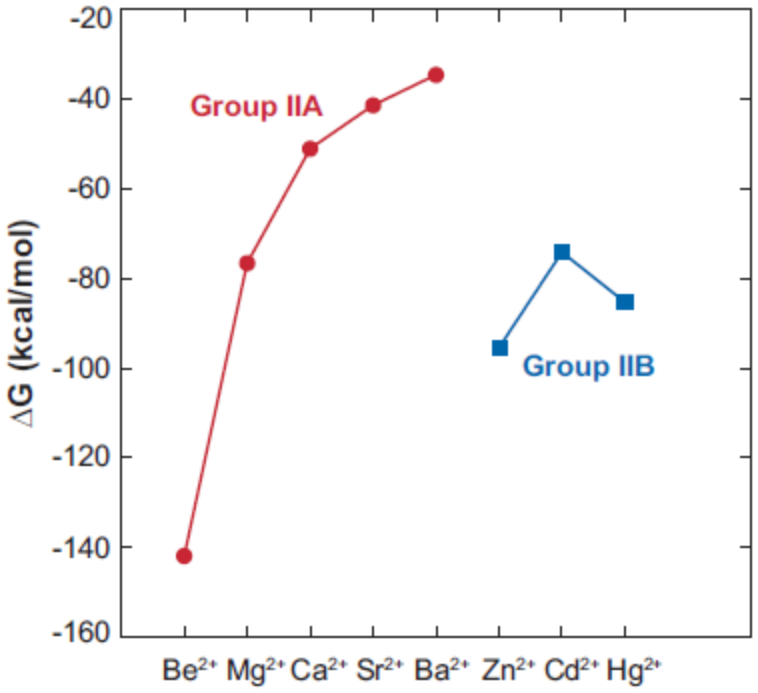


Figure 3.18: The binding gas-phase free energies of $M^{q+} + nH_2O \rightarrow [M(H_2O)_n]^{q+}$ calculated for IIA and IIB metals. Note Zn^{2+} has higher ΔG value than Mg^{2+} and Ca^{2+} *(164).*

In summary, the CD studies discussed above have laid ground work to the effort to refine the Zn^{2+} binding residues/region in the Protα. Protα(50-89)N50W similar to the full length protein undergoes a structural change in the presence of Zn^{2+}. The temperature dependent folding of the peptide both in the presence and absence of Zn^{2+} can be attributed to decrease in repulsive forces at higher temperature leading to folding of the peptide. By using the polyglutamic acid and scrambled Protα(50-89)N50W it was understood that Zn^{2+} binding and folding of the peptide is not solely due to electrostatic attractions between the cation and the peptide but there is sequence specificity between them. The selective interactions of the peptide with Zn^{2+} could be explained by the higher charge accepting ability of Zn^{2+} from the peptide than other divalent metals and stability of Zn^{2+} complexes.

Chapter 4

Zn^{2+} and temperature induced structural changes in Protα(50-89)N50W studied using Differential Scanning Calorimeter (DSC)

Section 4.1 Introduction to DSC

Background

Calorimetry is a primary technique for measuring the thermal properties of materials to establish a connection between temperature and specific physical properties of substances and is the only method for direct determination of the enthalpy associated with the process of interest *(119)*. Calorimeters are used frequently in chemistry *(120)*, biochemistry *(121)*, cell biology *(122)*, biotechnology *(123)*, pharmacology *(124)*, and recently in nanoscience *(125)* to measure thermodynamic properties of the biomolecules and nano-sized materials. Among the various types of calorimeters, differential scanning calorimeter (DSC) is a popular one. DSC is a thermal analysis apparatus measuring the amount of heat absorbed or released by the sample as a function of temperature and time. In other words, the device is a thermal analysis instrument that determines the temperature and heat flow associated with material transitions as a function of time and temperature *(126)*. During a change in temperature, DSC measures a heat quantity, which is radiated or absorbed excessively by the sample on the basis of a temperature difference between the sample and the reference material *(126)*.

Calorimetry is particularly applied to monitor the changes of phase transitions *(127)*. Thermal transition temperatures (Tt; melting points) of the samples are also determined in solution, solid, or mixed phases such as suspensions *(128-131)*. It is also used for the study of biochemical reactions, which is named as a single molecular transition of a molecule from one conformation

to another *(128)*. In a basic DSC experiment, heat energy is introduced simultaneously into a sample cell (which contains a solution with the molecule of interest) and a reference cell (containing only the solvent). Temperatures of both cells are raised identically over time. The difference in the input energy required to match the temperature of the sample to that of the reference would be the amount of excess heat absorbed or released by the molecule in the sample (during an endothermic or exothermic process, respectively) *(127-129)*. As a result of the presence of the molecule of interest, more or less energy is required to bring the sample to the same temperature as the reference; hence, the concept of heat excess comes into the picture (Figure 4.1).

As a powerful analytical tool, DSC is capable of elucidating the factors that contribute to the folding and stability of biomolecules. Changes in the capacity of a molecule to take up heat, commonly known as heat capacity (Cp) are believed to originate from the disruption of the forces stabilizing native protein structure. For example, this includes van der Waals, hydrophobic, and electrostatic interactions, hydrogen bonds, hydration of the exposed residues, conformational entropy, and the physical environment (such as pH, buffer, ionic strength, excipients) *(128)*. Therefore, thermodynamic parameters obtained from DSC experiments are quite sensitive to the structural state of the biomolecule. Any change in the conformation, like a mutation in the sequence would affect the position, sharpness, and shape of transition(s) in DSC scans.

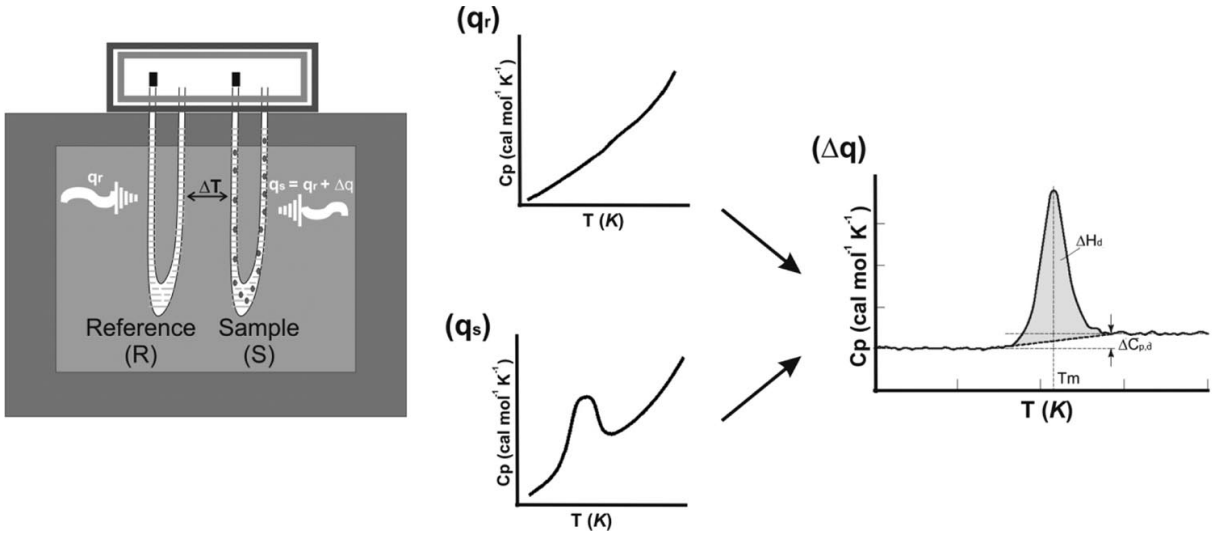


Figure 4.1: Experimental setup for a DSC experiment. The amount of heat required to increase the temperature by the same increment (ΔT) of a sample cell (qs) is higher than that required for the reference cell (qr) by the excess heat absorbed by the molecules in the sample (Δq). The resulting DSC scans with the reference subtracted from the sample show how this excess heat changes as a function of temperature T (*K*), Temperature, kelvin; ΔH, change in enthalpy; ΔCp,d, change in Cp; Tm, transition and melting point; d, denatured *(130)*.

Basics:

<u>Heat Capacity:</u>

As the name implies, heat capacity is a measure of the capacity of any object to take up heat energy. At the molecular level, this heat energy will be distributed among the available degrees of freedom and partitioned into kinetic energies (related to vibrational, rotational, or translational motions) or potential energies (related to changes in inter atomic potential energies, bond stretching, bending, breaking, and so forth). It follows that the more ways there are of distributing heat energy in a substance, the higher will be its heat capacity. This roughly explains

why liquid water has such a relatively high heat capacity; heat energy is diverted into breaking residual intermolecular hydrogen bonds rather than raising the temperature.

Heat capacity is central to fundamental thermodynamics because both the absolute enthalpy (H) and the entropy (S) of any object are integral functions of its heat capacity where Cp is the heat capacity at constant pressure, T is the absolute temperature, and the integral is taken from absolute zero (0 K) to the required temperature. H(0) and S(0) are the zero-point enthalpy and entropy, respectively.

$$H(T) = H(0) + \int Cp\, dT$$

$$S(T) = S(0) + \int \left(\frac{Cp}{T}\right) dT$$

Enthalpy:

The enthalpy of the endothermic or exothermic event is determined by the integration of the area under the DSC peak, which is often reported in kcal/mol *(131-134)*. The shape of the thermally induced event is described by the width of the transition at half height of the peak ($T_{1/2}$), whereby, the peak is defined by the difference between the lower (T_S) and upper boundaries (T_L) of the phase transition (Figure 4.2).

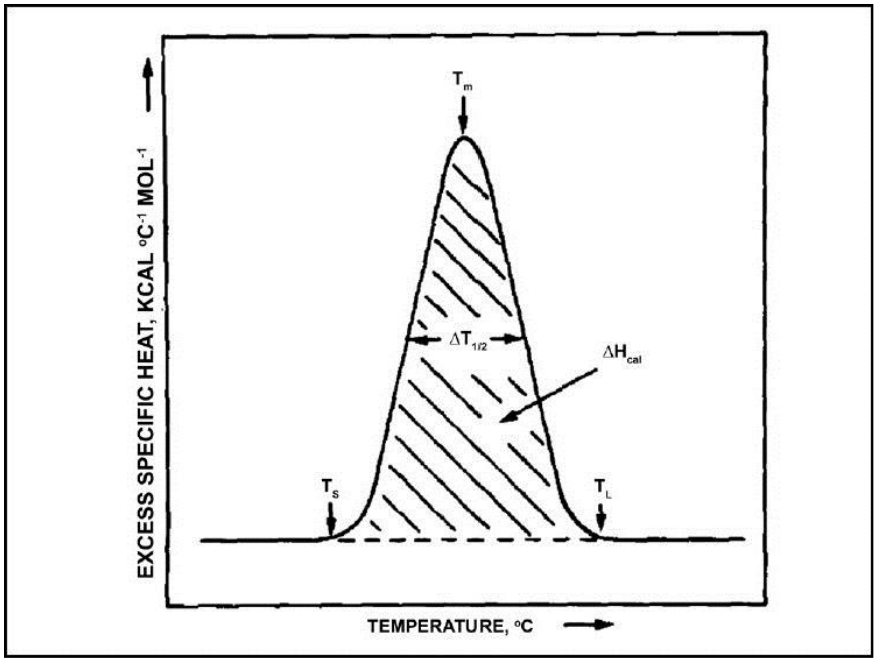


Figure 4.2: A sample endotherm showing Enthalpy, $T_{1/2}$ and T_m *(131)*.

Modulated differential scanning calorimetry (MDSC)

DSC measures the difference in heat flow between a sample and reference as both are subjected to a linear change in temperature. The heat flow signal is a combination of 'kinetic' (f(T,t)) and 'heat capacity' (Cp•dT/dt) responses, i.e. the heat flow from the sample (or reference) is a function of both heat capacity of the sample and of any kinetic heat flow component *(135)*. The heat flow signal is given as

$$\frac{dQ}{dt} = Cp.\frac{dT}{dt} + f(T,t)$$

where Q is the (heat) energy, Cp is the 'thermodynamic' heat capacity (i.e. that due to the energy stored in vibrations, rotations and translations of molecular constituents of the sample), T is the

absolute temperature and f(t,T) is some function of the time and temperature which expresses the calorimetric response of any kinetically controlled chemical or physical phenomena.

Modulated DSC measures both kinetic and heat capacity in a single experiment by superimposing a modulated heating rate on top of a linear heating rate (Figure 4.3). Advantages of MDSC include high resolution, high sensitivity, superior sensitivity exhibited by MTDSC to measure glass transitions.

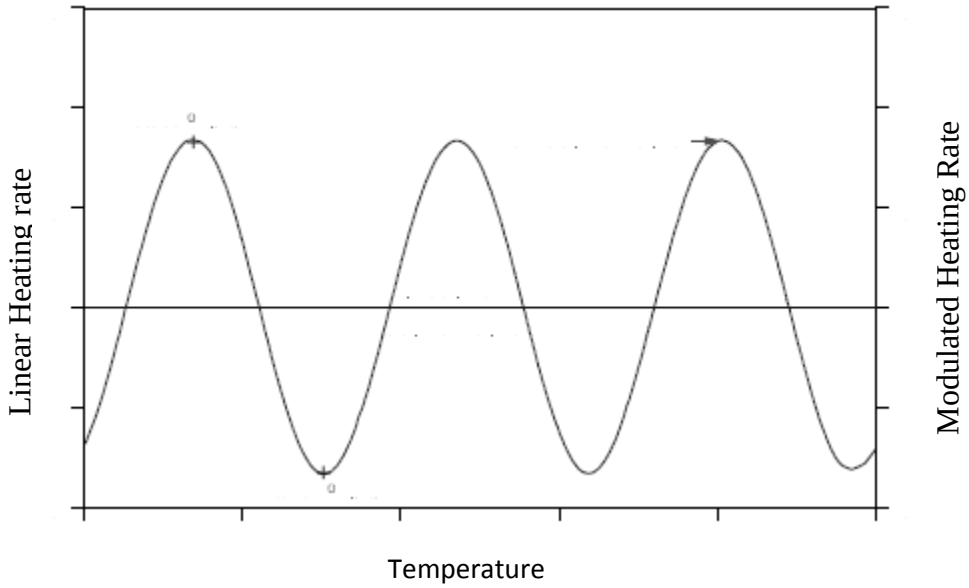


Figure 4.3: MDSC temperature profile. The linear change in temperature permits the measurement of heat flow while the modulated change permits the calculation of heat capacity.

Applications

Considering the ability to measure enthalpy changes and phase transitions, there are multiple applications for a versatile tool like DSC. There are several reviews on its application to various fields like proteins*(136)*, protein of pharmaceutical interest *(137)*, protein mutations *(138)*, protein-

ligand interactions *(139,140)*, protein folding *(141)*, nucleotides *(138)*, other macromolecules *(142)*, lipids *(142-144)*, drug-lipid interactions *(143)*, and protein-lipid interactions *(144)*.

DSC in Proteins

One of the earliest DSC applications was to study thermally induced, cooperative conformational changes of small proteins. Protein thermodynamics, during unfolding, is measured as an enthalpy change as a function of temperature. The temperature at the maximum point of the Cp curve (Tm) represents the macromolecule stability *(145-147)*. DSC based thermal analysis provides insight into protein unfolding and forces involved in conformational stability. Protein denaturation process is highly cooperative with disruption of many forces and bonds, including hydrogen bonds, hydrophobic interactions, and many non-covalent interactions *(148)*. DSC allows for the direct study of thermal stability, over a very large concentration range, in the absence of light, thus photosensitive proteins such as bovine lens crystallins can be analyzed *(148)*.

A typical DSC data for the heat capacity increment observed upon thermal unfolding of a globular protein in aqueous solution is shown below (Figure 4.4) *(129)*. The thermal unfolding of globular proteins in water is a classic example, where direct calorimetric (DSC) measurements, supported by more indirect methods, show that an overall heat capacity increment (positive *C*p) is an almost universal characteristic of the unfolding transition *(119, 138)*. The protein shows an unfolding endothermic event as the protein unfolds and as the heat energy being applied disrupts intramolecular interactions. As the temperature was increased the protein moved from a folded state to an unfolded state. The difference in the baseline before and after the melting temperature is measured as the heat capacity (Cp). The heat capacity of the unfolded peptide is significantly greater than that of more compact native conformation.

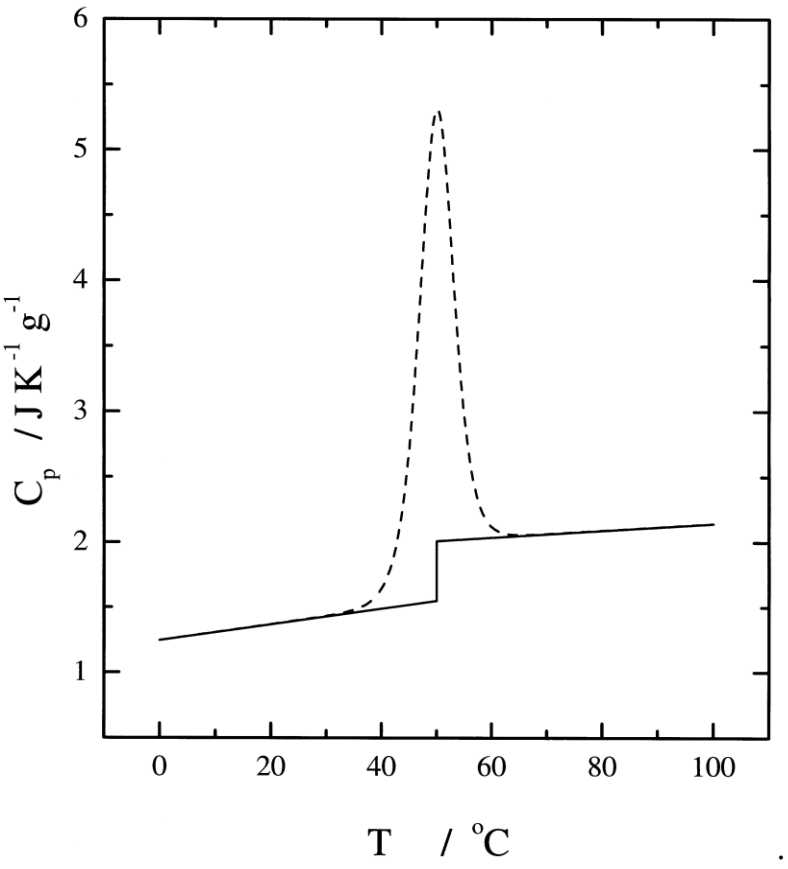


Figure 4.4: Typical DSC data for the heat capacity increment observed upon thermal unfolding of a globular protein in aqueous solution. The dotted line shows the overall heat energy uptake associated with the actual unfolding transition. The difference in the baseline is measure as heat capacity (Cp) *(129)*.

Protein Heat Capacity:

The measurement of heat capacities (Cp) and heat capacity changes (ΔCp) for proteins in dilute (<1 mg mL^{-1}) aqueous solutions is now relatively straightforward using commercially available differential scanning calorimeters (DSC) specifically designed for the purpose, most usually in the 0-110 ^{o}C temperature range *(119, 149)*. Figure 4.5 shows "Typical" DSC data for the thermal

unfolding of a simple globular protein in solution at varying pH. The Figure illustrates various features that will be considered.

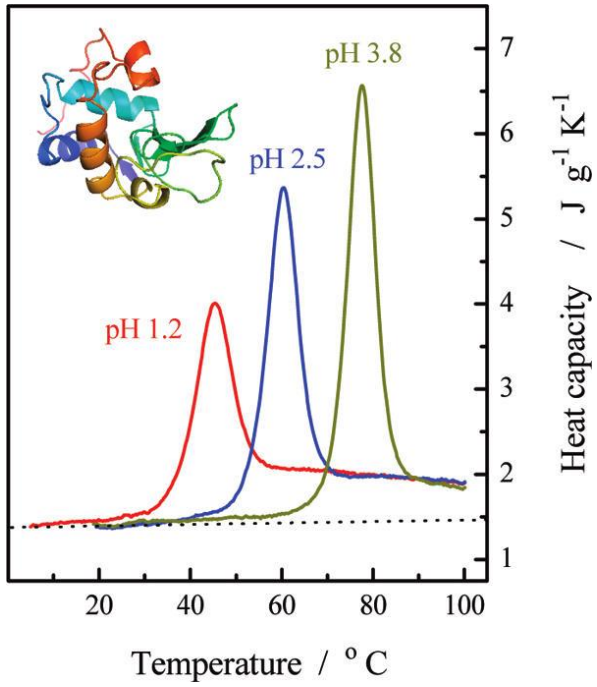


Figure 4.5: Typical DSC data for the unfolding of a small globular protein (lysozyme) in solution at various pH values. The increase in area under each endotherm with higher Tm and the higher heat capacity baselines after the unfolding transitions are both indications of the significant positive ΔCp commonly associated with such processes. *(127)*

For well-behaved proteins undergoing reversible thermal transitions, the heat capacity curves obtained by DSC (Figure 4.5) show several characteristic features. Starting from low temperature, the apparent protein heat capacity gradually increases until the onset of thermal unfolding, accompanied by a sharp peak in Cp corresponding to an endothermic unfolding transition. The peak of the transition gives the midpoint temperature (Tm), and the area under the transition gives the enthalpy of unfolding (ΔHunf) at this temperature. The heat capacity of the protein above the transition, once it has returned to baseline, is typically higher than would be

anticipated from extrapolation of the low-temperature baseline. Consequently, the thermal unfolding shows a positive heat capacity change, ΔCp. This is also indicated by the marked increase in the area under the transition peak (ΔHunf) with increasing Tm.

The sort of behavior shown here is usually only seen for relatively small proteins undergoing reversible unfolding, typically at low pH and low concentration. Unfortunately, most proteins do not behave that way, especially near neutral pH. Denaturation of unfolded protein is frequently accompanied by aggregation or other irreversible effects *(150)*. The DSC study on Protα(50-89)N50W would set the stage for introducing a new alternative method to study conformational changes in IDPs in the presence and absence of metal cations.

Section 4.2 Research Strategy

Protα lacks a native tertiary or secondary structure according to the usual criteria provided by NMR, CD, and SAXS *(43, 60)*. To compliment the CD data the Protα(50-89)N50W peptide has been studied using Modulated Differential Scanning Calorimetry (MDSC).

Specific Aim: The goal of this study was to investigate the conformational stability of Protα(50-89)N50W during the process of thermal denaturation in the presence and absence of zinc ions, Zn (II), and to estimate the thermodynamic parameters of protein unfolding and zinc binding with the purpose of obtaining a better understanding of the Protα(50-89)N50W/zinc interaction.

Based on temperature dependent CD studies we hypothesize the peptide undergoes an unfolded to folded transition upon heating in DSC. Our first aim was to look at the conformational changes of the peptide itself (metal free) upon heating. The subsequent set of experiments was carried out by the addition of 50 eq. of Zn^{2+} to the peptide.

Experiments:

All DSC measurements were carried out on TA Instruments Q2000 calorimeter. An optimum concentration of the peptide that gives a good signal on DSC has been determined to be 50 mg/ml. The empty sample pans were weighed and a 10 µl sample of Protα(50-89)N50W (50 mg/ml) dissolved in 50 mM phosphate buffer at pH 7.2 was placed in a hermetically sealed t-zero aluminium pans. N_2 is used as a purge gas and is preheated for 30 mins to ensure uniform temperature environment. DSC measurements were performed by placing the peptide in the sample pan and buffer in the reference cell at a scan rate of 1 ^{o}C/min from 25 to 80 ^{o}C under modulated conditions. 50 mM phosphate buffer at pH 7.2 was used as buffer. Thermograms of

the peptide with and without Zn^{2+} in the sample were obtained. The thermodynamic parameters like enthalpy, melting temperature, and heat capacity were obtained.

Section 4.3 DSC Measurements

1.3 Investigation of the structural changes in the Protα(50-89)N50W upon thermal denaturation using differential scanning calorimetry (DSC)

MDSC thermogram of prothymosin-α (50-89 N50W) in 50 mM phosphate buffer was recorded at pH 7.5. The temperature was raised from 25–80^{o}C. The concentration of the peptide used was 50 mg/ml. The peptide exhibits an exothermic event indicating collapse into a more compact form and that the new structure is forming intramolecular interactions. Hence, heat is given off as new intramolecular interactions are formed (Figure 4.6).

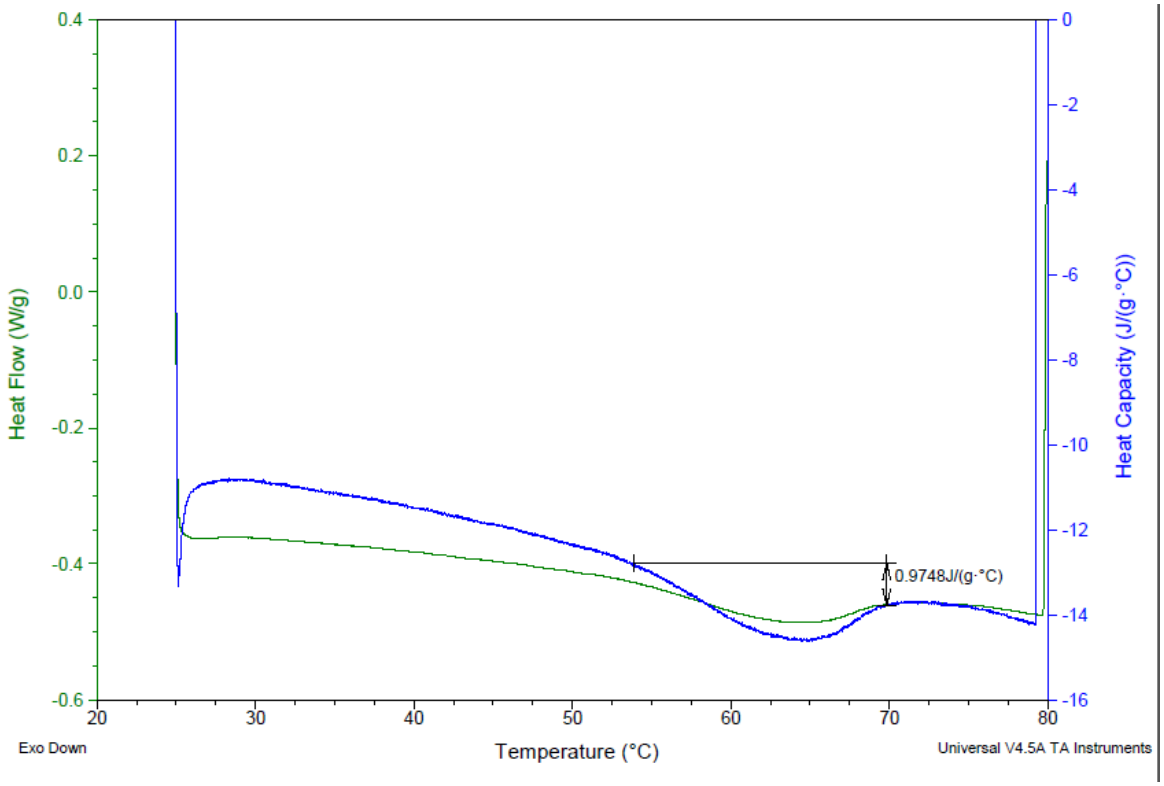


Figure 4.6: The thermal induced folding of Protα(50-89)N50W peptide monitored by DSC under modulated conditions. Blue and green traces indicate heat capacity and heat flow respectively. The heat capacity of the peptide folding is 0.9748 J/(g*(^{o}C)).

With the addition of 50 eq. Zn^{2+} the peptide continued to show an exothermic event which can be interpreted as a folding event. However, the transition is broadened in the Zn^{2+}-loaded peptide suggesting a smaller, temperature induced structural rearrangement. Figure 4.7 shows the thermogram for Zn^{2+}-loaded peptide.

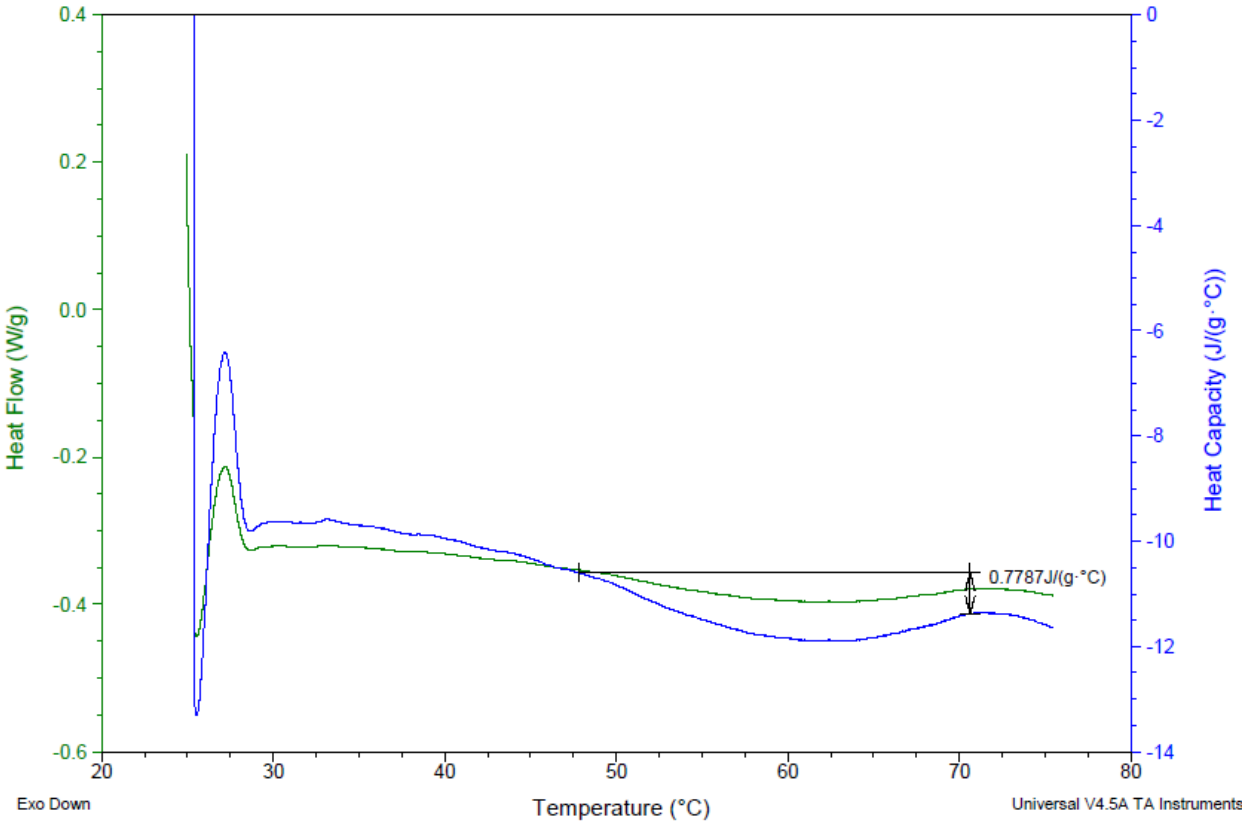


Figure 4.7: The thermal induced folding of 50 eq. Zn^{2+}-loaded Protα(50-89)N50W peptide monitored by DSC under modulated conditions. Blue and green traces indicate heat capacity and heat flow respectively. The heat capacity of the peptide folding is 0.7787 J/(g*^{o}C).

The mid transition temperature, enthalpy, and specific heat of the peptide before and after Zn^{2+} addition are listed in the table below. These results will be addressed in the discussion section below.

Sample Name	Mid-transition Temperature (^{o}C)	ΔH (kJ/mol)	ΔC_p (J/g ^{o}C)
Protα(50-89)N50W	*63.66*	-68.7	-0.9748
Protα(50-89)N50W with 50 eq. of Zn^{2+}	*60.63*	-50.2	-0.7787

Table 4.1: The thermodynamic parameters: mid-transition temperature, enthalpy, and heat capacity for the peptide both in the presence and absence of Zn^{2+}

Discussion:

The temperature dependent profile of Protα(50-89)N50W in the presence and absence of Zn^{2+} was studied using DSC. The peptide showed an exothermic event in the presence and absence of Zn^{2+}. Literature on DSC studies of folded proteins has explained an endothermic event as a defining feature of protein unfolding in solution resulting in a positive ΔCp and an exothermic event as protein aggregation resulting in low heat capacity than the folded form of the protein (Figure 4.8) *(152).*

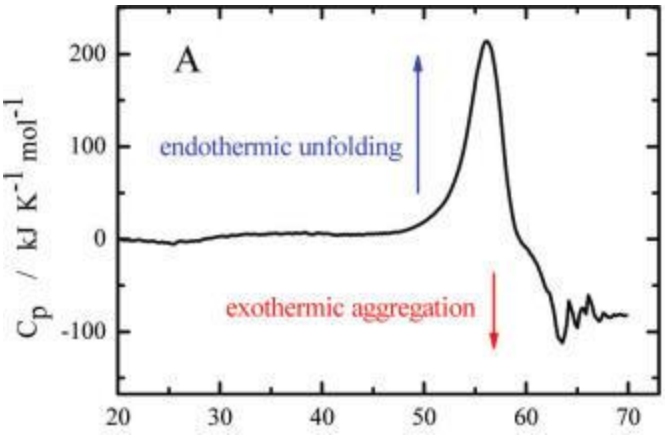


Figure 4.8: DSC data for thermal denaturation of yeast phosphoglycerate kinase, illustrating exothermic baseline distortion and heat capacity decrease caused by irreversible precipitation of unfolded protein *(152).*

However we strongly believe that the exothermic event seen in Protα(50-89)N50W peptide both with and without Zn^{2+} in the sample does not correspond to protein aggregation because of three reasons. The first is the temperature dependent CD runs on the peptide showed a reversible event indicating no aggregation of the peptide upon heating and cooling. The second point of support is the Cp value. The Cp value of an aggregated peptide shows a much lower value after heating compared to the folded peptide (Figure 4.9). In the case of Protα(50-89)N50W peptide an extrapolated line shows a continuous trend, whereas there is an abrupt and large change in Cp upon aggregation (note discontinuity of the base line in Figure 4.9). Lastly, given the high negative charge of the peptide aggregation is highly unlikely.

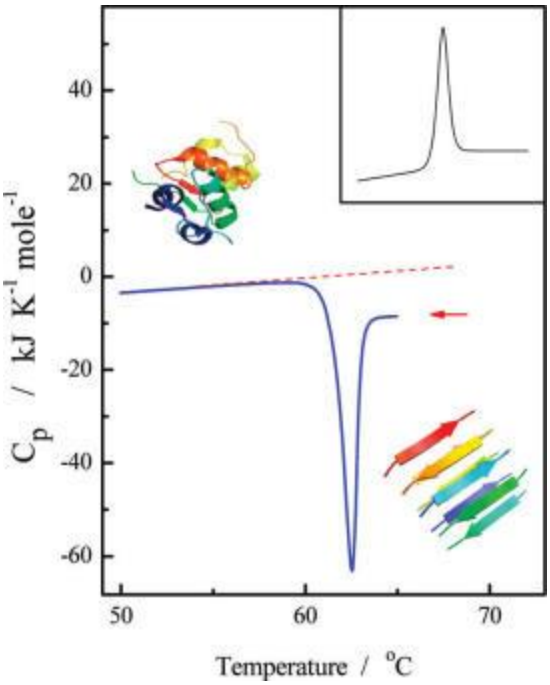


Figure 4.9: DSC thermogram of the thermal aggregation of insulin in solution. The inset shows the more typical thermogram seen for protein unfolding (without aggregation). Note the decreased heat capacity for the fibril state (arrow) *(152)*.

The exothermic event in the peptide can be explained as the Protα(50-89)N50W peptide behaving opposite way a folded protein behaves. That is an unfolded protein unlike the folded protein collapses and adopts a structure upon heating. The DSC data of Protα showed an exothermic event indicating protein folding both in the presence and absence of Zn^{2+}. These results coincide well with the CD temperature studies. In temperature dependent CD studies of the peptide the shape of the spectra is congruous with increasing structure upon heating from 20-80 ^{o}C both in the presence and absence of Zn^{2+}.

In the presence of Zn^{2+} the folding event is broadened and the ΔH value is less negative indicating decreased cooperative folding in the protein compared to the no Zn^{2+} sample. A decreased folding of the Zn^{2+} loaded peptide can be explained by decreased co-operativity

associated with the exothermic event implying that Zn^{2+} disrupts or interferes with the natural folding/organization of the peptide when heated. This may be due to the fact that Zn^{2+} is now bound to the peptide and occupying sites on peptide backbone that are important during peptide folding by heat or Zn^{2+} pre-organizes the peptide and the folding has taken place simply due to the addition of Zn^{2+} so there is less folding that can occur upon heating when compared to the apo-peptide. Additional evidence is seen when the enthalpy associated with this event is examined. A decreasing enthalpy indicates that there may be less intra or inter molecular interactions when Zn^{2+} is present *(151)*.

One would expect that the peptide folding is indicative of buried hydrophobic surfaces. However, for acidic peptides like a Protα(50-89)N50W a negative correlation exists between the buried surface area and the combination of net high charge and low hydrophobicity *(16, 116, 151)*. If the folding in the peptide is real, the deviation of the peptide from paradox (unfolding upon heating is the regular trend) needs a deeper insight.

The literature available on DSC of proteins so far has not explained a negative ΔCp value. The exothermic event has been treated only as protein aggregation. A literature study using DSC on a family of natively disordered fragments from Escherichia coli thioredoxin has reported heat capacity analysis of fragments of thioredoxin protein. Most of the fragments analyzed showed a positive ΔCp indicating that the fragment behaves like a globular soluble protein, which buries its hydrophobic patches to a greater extent than the polar ones. However, a fragment (32-108) showed negative ΔCp values throughout the whole temperature range and for fragment 51-108 the ΔCp becomes negative beyond 66 ^{o}C. The study has left the negative ΔCp value seen for fragment 32-108 unexplained.

To conclude spin lattice relaxation (T_1) measurements at different temperatures on NMR and a means of digging deeper into analysis of temperature dependent profile of the peptide is essential to obtain any further information about temperature dependent folding of Protα.

Chapter 5

Spin-lattice relaxation measurements of EDTA and Zn^{2+}

Section 5.1 Introduction on NMR Relaxation

NMR is one of the most common methods used for the biophysical characterization of protein-metal interactions *(152)*. The first step to obtain structural information by NMR measurements is the assignment of resonances. In the case of IDPs like Protα this can be challenging due to extensive signal-overlap and exchange broadening of amide proton signals. For structured proteins, resonance assignment is done with a set of spectra, typically ^{15}N, and ^{1}H resolved triple resonance experiments that contain information linking ^{1}HN, $^{13}C\alpha$, $^{13}C\beta$, $^{13}C\gamma$, and ^{15}N chemical shifts. Figure 5.1 shows sample ^{1}H-^{15}N HSQC spectra. The figure shows how a well resolved spectrum (on the left) of a folded protein differs from a signal overlapped spectrum (on the right) of the same protein in a denatured state. The sequence homogeneity in IDPs results in chemical shift degeneracy and signal overlap demanding a different approach. A measure of relaxation times is a possibly method of approach to study dynamics and metal binding in IDP's. A detailed strategy on how this type of study can be applied to Protα(50-89)N50W will be discussed in chapter 6.

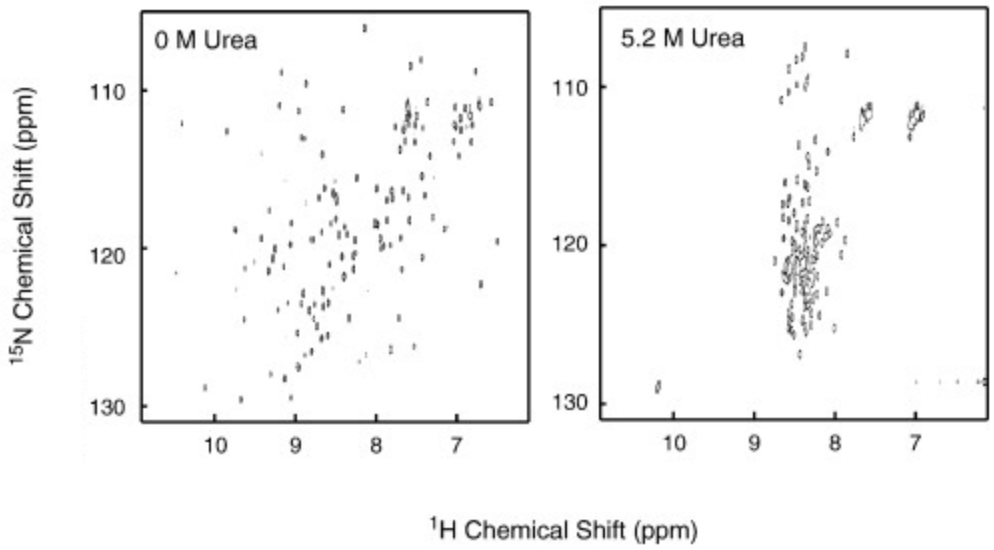


Figure 5.1: $^{1}H–^{15}N$ HSQC spectra of uniformly ^{2}H-^{13}C-^{15}N labeled WT-Interleukin-1 showing signal dispersion on the left and overlap on the right upon denaturation with 5.2 M urea *(153)*.

Relaxation measurements:

Relaxation of a nucleus is defined as a process by which excited nuclear magnetization vector of the selected nucleus loses the extra energy and returns to equilibrium. There are two types of relaxations; spin-lattice (T_1 or longitudinal relaxation or relaxation in Z-axis) and spin-spin (T_2 or transverse relaxation or relaxation in x-y plane). T_1 relaxation corresponds to the process of establishing the normal Gaussian population distribution of α and β spin states in the magnetic field. T_2 is loss of phase coherence among nuclei. T_2 is less than or equal to T_1 since return of magnetization to the z-direction inherently causes loss of magnetization in the x-y plane. Relaxation rate is inverse of relaxation time.

$$R = 1/T$$

R is relaxation rate, T is relaxation time.

Spin-lattice relaxation

When molecules absorb electromagnetic radiation they are transferred from lower energy states to higher energy states. This leaves the system out of equilibrium, and the system must undergo transitions from the upper states to lower states to get back to equilibrium. This process is illustrated in Figure 5.2.

For protons, the energy level diagram has two energy levels, the lower state with protons "spin up", or aligned with the external field, and the higher state with protons "spin down", or aligned against the external field. Electromagnetic radiation in the radiofrequency region of the spectrum causes protons to "flip", or jump to higher energy levels. The process whereby the system returns to equilibrium involves spin flips to return to the lower energy state, and the excess energy is lost to the surroundings in the form of heat. The surroundings in magnetic resonance experiments are called the "lattice", therefore the name spin-lattice relaxation. The characteristic life-time of a spin in the upper state is called the spin-lattice relaxation time T_1. T_1 is the average length of time that a proton remains in the excited state or higher energy level. The spin-lattice relaxation time is also called the longitudinal relaxation time.

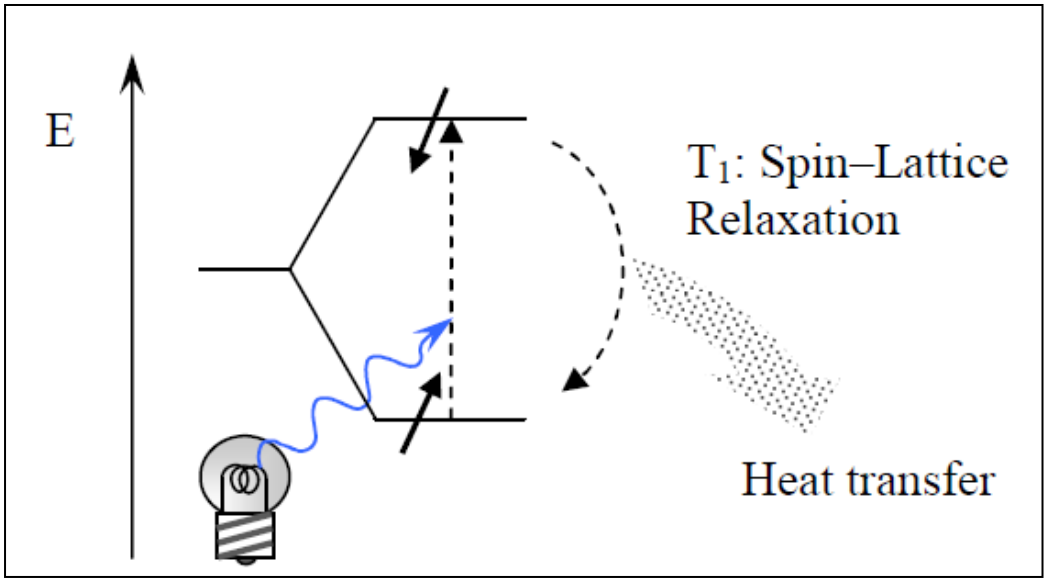


Figure 5.2: Basic representation of absorption and spin-lattice relaxation for spin I=1/2 system

An easy way to visualize the process is shown in Figure 5.3, which illustrates an older method for measuring T_1's called "saturation recovery." At left the spins start at equilibrium. The radiofrequency source is then turned on, which causes transitions between the two levels. After the source has been on for a short time the populations (number of spins in a given level) of the two levels are equalized. With equal populations, the two levels are said to be saturated and will no longer give a resonance signal. The source is then turned off, which allows the nuclei to return to equilibrium some time later. A good rule of thumb is that it takes about $5xT_1$ for the system to return to equilibrium *(154)*.

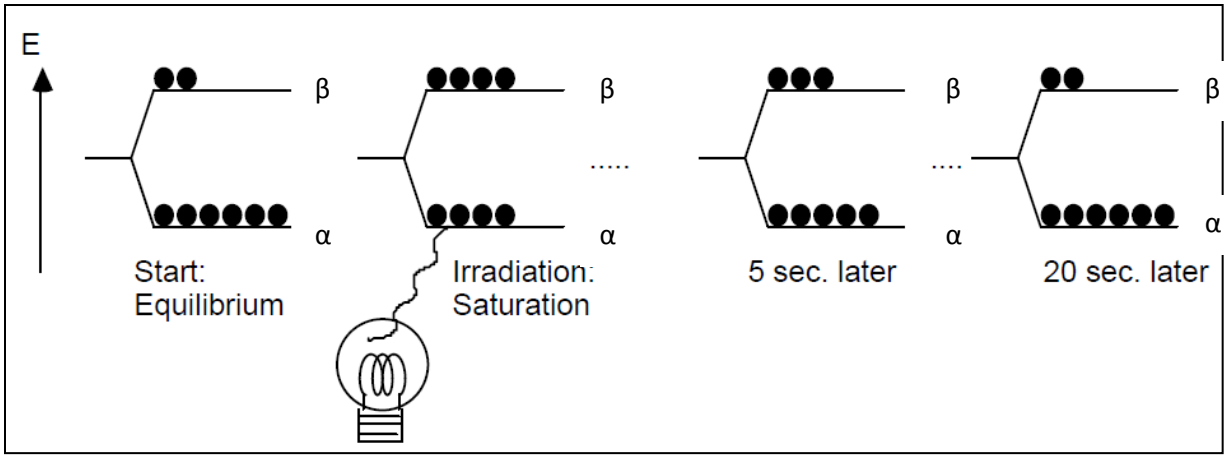


Figure 5.3: Saturation recovery method for spin-lattice (T_1) measurement. Irradiation leads to saturation and after the radiation source is turned off the system begins to return to equilibrium.

The population ratio in the α and β spin states is given by the Boltzmann equation that relates the population ratio to the energy difference between the two states (ΔE) and the absolute temperature *(154)*.

$$N_\beta / N_\alpha = e^{-\Delta E/(K_B.T)} \qquad (1)$$

Where N_β and N_α stand for population of nuclear spins in α and β states, ΔE is the energy difference between the two spin states, K_B is the Boltzmann constant, and T is the temperature in Kelvin.

The NMR signal is caused by the net magnetization of the spins M_z, which is proportional to the difference in the populations of the two levels:

M_z = (# spins in lower state) - (# spins in upper state)

Spin-lattice relaxation is a first order kinetic process given by Bloch equation:

$$\frac{dM}{dt} = -\frac{1}{T_1}(M_z - M_o) \qquad (2)$$

Where M_z is the magnetization at time t and M_o is the magnetization at equilibrium. Integrating equation 2 for the recovery from saturation gives:

$$M_z = M_0 e^{-t/T_1} \quad (3)$$

Spin-Spin relaxation

The spin–spin relaxation is the mechanism by which the transverse component of the magnetization vector, exponentially decays towards its equilibrium value of zero in nuclear magnetic resonance (NMR) and magnetic resonance imaging (MRI).

Spin-spin lattice is not discussed in detail as the data focuses on T_1 measurements and T_2 will not influence this study.

Measurement of T_1 by spin-inversion recovery

For T_1 relaxation to occur there must be magnetic field fluctuations in the x,y direction. Such fluctuations are most effective when they occur at the Larmor frequency (ν_o). T_1 relaxation is thus field dependent, since ν_o varies with the field. The principal source of fluctuating magnetic fields in most molecules is molecular motion. The motion of a molecule can be characterized by a correlation time, τ_c (s). The correlation time is the average time that a molecule spends in a given orientation *(154)*. Large molecules tumble slowly and therefore have long τ_c. Small molecules tumble very rapidly and therefore have short τ_c.

The step by step procedure of how T1 relaxation is measured (Figure 5.3) by spin inversion recovery experiment will be discussed in detail. When a system is in equilibrium it has more spins in α state than in β state. When a pulse sequence (180^o-t-90^o) is applied to the nuclei it causes inversion of spins in the α and β states (see top of Figure 5.4 for visual representation of this pulse sequence). This creates a population inversion where there are more spins in the higher state than in the lower state.

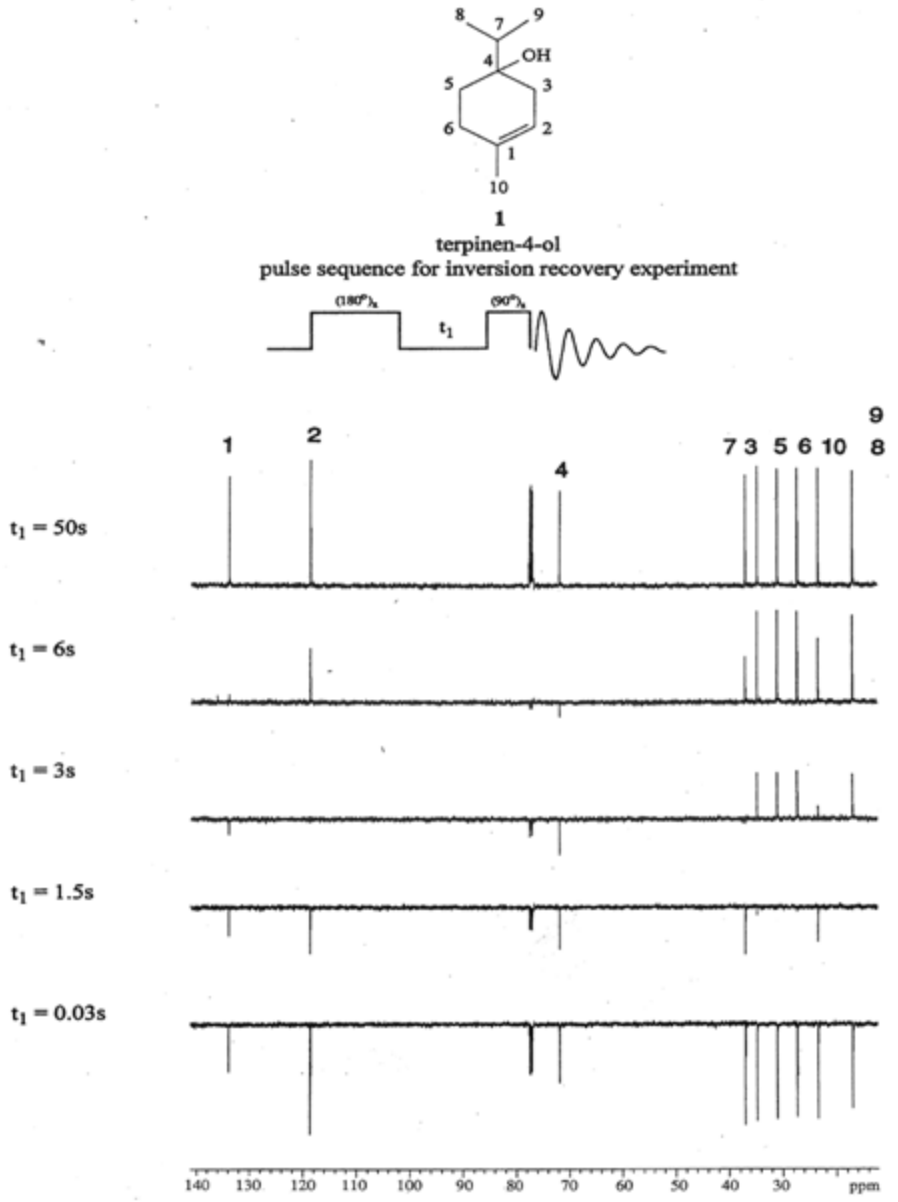


Figure 5.4: ^{13}C NMR inversion recovery experiment on terpinen-4-ol *(155).*

The spectrum of the system in the inverted state contains all negative peaks (shown at t_1 = 0.03s in the Figure 5.4). For short delayed times (t_1) values, the population remains inverted and peaks in the spectrum remain negative. For longer delayed times, the system reaches equilibrium and the spectrum starts to show positive peaks. At even longer delay times the system has reached equilibrium prior to the second pulse and typical NMR spectrum is recorded. The spectra for

various delayed times are plotted on the same chart, giving the so-called partially relaxed spectra for the system, as shown in Figure 5.5. Note that the different carbons in the molecule relax at different rates. This is due to their local environment and molecular motion. The point where a peak changes from negative to positive is a rough measure of the T_1, and $t = T_1 \ln 2$.

To be more appropriate the T_1 value for each chemical environment can be obtained by fitting the peak intensity of each peak to the equation:

$$\ln(M_z - M_o) = \ln(2M_0) - \frac{\tau}{T_1} \qquad (4)$$

Notice that taking the natural logarithm (ln) of both sides of the above equation would give $\ln(M_z) = \ln(M_o) - t/T_1$, the difference here is that the signal for a small t is negative.

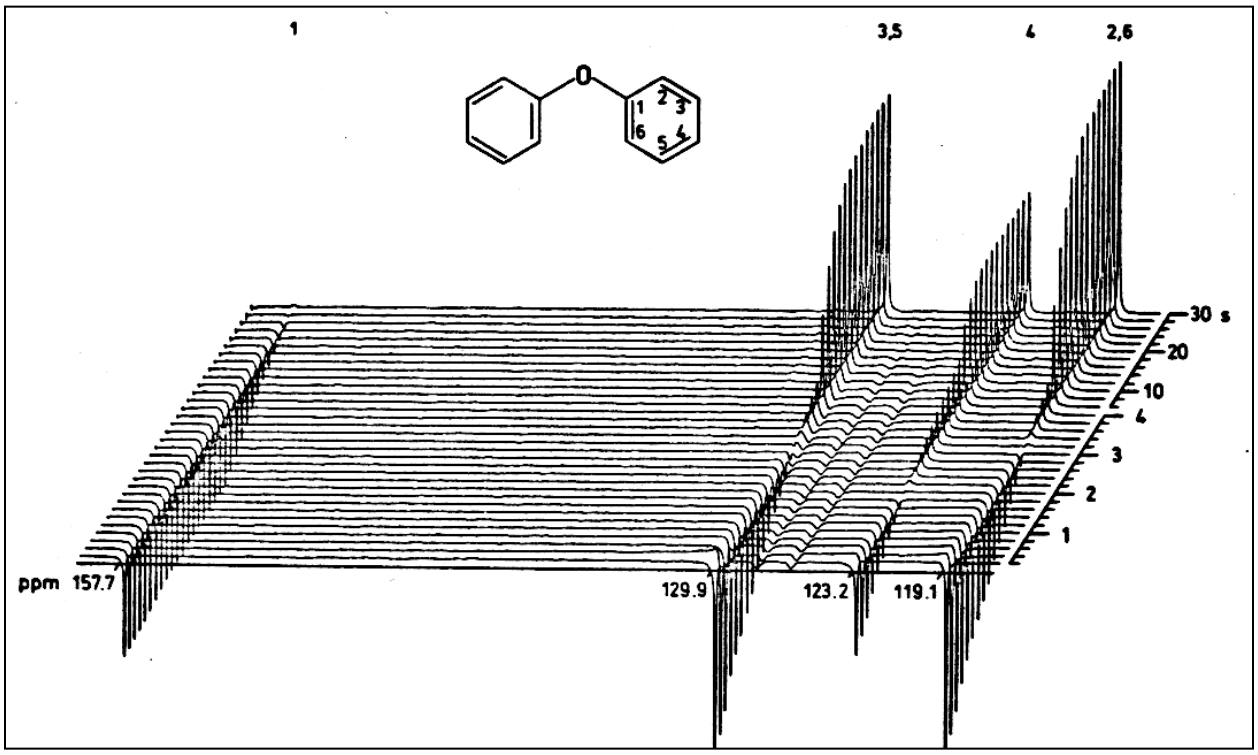


Figure 5.5: Measurement of T_1 by inversion recovery of diphenyl ether *(155)*.

There are several mechanisms by which molecular motions can influence nuclear relaxation *(154)*: direct dipole-dipole interactions with nearby magnetic nuclei (DD), chemical shift anisotropy effects (CSA), quadrupole-electric field gradient interaction (QR) and rapid modulation of *J*-coupling (SC). In addition to molecular motion, rotational transitions can also be the source of fluctuating magnetic fields (SR).

Dipole-Dipole Relaxation (DD)

The direct dipole-dipole coupling interaction is very large (often kilohertz) and depends on the strength of the dipolar coupling, on the orientation/distance between the interacting nuclei and on the motion. The equation below accounts for the relaxation due to the dipole-dipole interactions.

$$R_1 = \left(\frac{h}{2\Pi}\right)^2 . \ \gamma_H^2 . \gamma_X^2 . (r)^{-6} . \tau_c \qquad (5)$$

Where h is Planck's constant, γ_H and γ_X are the gyromagnetic ratios of proton and any other nucleus being relaxed, r is the distance between proton and other nucleus being relaxed, τ_c is the correlation time. This coupling is not seen in non-viscous solutions because it is averaged to zero by tumbling of the molecule. However, as the molecule tumbles in solution the dipole-dipole coupling is constantly changing as the vector relationships change. This creates a fluctuating magnetic field at each nucleus. To the extent that these fluctuations occur at the Larmor precession frequency, they can cause nuclear relaxation. This relaxation mechanism is particularly important for molecules containing protons because proton has the highest gyromagnetic moment of common nuclei; it is the most effective nucleus for causing DD relaxation. DD relaxation is the principal relaxation pathway for carbons with directly attached protons.

DD relaxation shows very strong distance dependence, and operates most effectively between directly bonded nuclei. At the usual spectrometer frequencies, small molecules (MW <1000) are tumbling too fast for the most effective relaxation (τ_c is too short). Thus the more rapidly a molecule or part of a molecule tumbles, the less effective DD relaxation is, and the longer T_1 becomes. The motion of large molecules (e.g. proteins) is slower compared to small molecules (τ_c is too long), and have the opposite relationship between molecular motion and T_1 (i.e., relaxation is more effective when the molecule moves faster). For ^{13}C NMR, dipole relaxation by directly attached protons (if any are present) is the principal relaxation mechanism. Thus quaternary carbons will have long relaxation times (hence their low intensity under normal conditions of spectrum acquisition), methine (CH) groups will have shorter relaxation times, and methylene (CH_2) groups still shorter, by approximately a factor of two. Methyl (CH_3) groups show highly variable T_1 values but can be very short as seen in Figure 5.4. They usually move much more rapidly than the rest of the molecule, and thus undergo less effective DD relaxation, but this may be complicated by SR relaxation.

Chemical Shift Anisotropy (CSA)

The chemical shift of a nucleus is a function of the orientation of the molecule with respect to the magnetic field (i.e. it is a tensor quantity), provided the nucleus is not at the center of tetrahedral or octahedral symmetry. As the molecule tumbles in solution, the chemical shift (and hence the magnetic field at the nucleus) is constantly changing and causes relaxation of the nucleus. The relaxation rate is proportional to the square of the gyromagnetic ratio and of the magnetic field strength, as well to the chemical shift anisotropy. The equation below applies to cylindrically symmetric molecules

$$R_1(CSA) = \frac{1}{T_1(CSA)} = \frac{2}{15}\ \gamma^2\ B_o^2(\sigma - \sigma_o)\tau_c \qquad (6)$$

B is the magnetic field, γ is the gyromagentic ratio, τ_c is correlation time, σ is chemical shift anisotropy (CSA) (if symmetrical CSA = 0).

The mechanism is never seen for protons, and is seen for carbon only when there are no attached protons (e.g., carbonyl compounds other than aldehydes and formates). Anisotropy is a property of p-orbitals. An s-orbital is isotropic hence why protons cannot access this mechanism of relaxation. Nuclei such as ^{199}Hg, ^{119}Sn and ^{77}Se with large chemical shift ranges are relaxed largely or entirely by the CSA mechanism. Operation of this mechanism causes reduction or loss of NOE effects.

Quadrupole Relaxation (QR)

This mechanism operates for spin >½ nuclei only, and only for nuclei which are not at the center of tetrahedral or octahedral symmetry. Spin ½ nuclei can be considered to have spherical charge distributions, but for spin >½ nuclei the charge distribution has the shape of an oblate or prolate spheroid. Electric field gradients in such molecules exert a torque on the quadrupolar nuclei. Tumbling of the molecule can then initiate transitions among the spin states (there is "friction" between the nucleus and the surrounding electrons, the "quadrupole coupling constant". The effectiveness of this relaxation mechanism is critically dependent on this coupling. If Q is small (as for ^{2}H and ^{6}Li) the nucleus behaves like a spin ½ nucleus, if it is large the nucleus can have very short T_1, and observation can be very difficult.

Spin Rotation (SR).

A local magnetic field is generated by the circular motion of electrons in a rapidly rotating molecule (or part of a molecule, such as a methyl group). The magnitude of this field changes

when the rotational energy levels change as a result of molecular collisions. These changes, if they occur at the Larmor precession frequency, can cause relaxation of nearby nuclei. The correlation time for SR relaxation is not molecular motion, but the lifetime of rotational quantum levels. Small molecules or freely spinning portions of larger molecules without other efficient relaxation mechanisms undergo SR relaxation. Better SR relaxation (shorter T_1) occurs at higher temperatures. The SR mechanism has been detected for methyl groups, and for quaternary carbons of small molecules which do not have other effective relaxation mechanism. Operation of this mechanism causes reduction or loss of NOE effects.

Nuclear Overhauser Effect (NOE)

When a proton is close in space to another proton (or any other nuclei with spin > 0), their magnetic dipoles interact. This interaction is distinct from *J* coupling, which is not a through space effect, but is mediated by polarization of bonding electrons in the molecule. The NOE effect is the change in spin state population of one proton (or other nucleus) when another magnetic nucleus close in space is irradiated. NOE is a consequence of DD relaxation where one nucleus is causing T_1 relaxation in the other and can be used to determine intra molecular distances.

The effects of these different relaxation mechanisms will be observed and reported in Zn^{2+}-EDTA studies.

Section 5.2 Research Strategy

The main question to be addressed by the proposed studies regarding zinc (Zn^{2+}) binding to the protein (Protα) is, if there is a single localized binding site or are there several degenerate binding sites? This might provide insight into how intrinsically disordered proteins that lack the most common ligands cysteine and histidine, bind Zn^{2+} and helps in the determination of zinc-binding residues in Protα. The results from NMR might also provide insight into CD and DSC results.

2D NMR experiments have revealed that the central segment of Protα is the Zn^{2+} binding region *(60)*. However because of the lack of heterogeneity in the central segment of the sequence the study could only determine the general metal binding region and not the specific metal binding sites/residues in the protein.

The fact that Protα lacks the most common Zn^{2+} binding residues cysteine and histidine the next common ligands for Zn^{2+} in Protα are carboxylate side chains of aspartic and glutamic acid. Relaxation measurements with selective ^{13}C isotope labeling of ^{13}CO groups in aspartic and glutamic acid is our method of choice to investigate Zn^{2+} binding in Protα.

^{13}C relaxation offers a number of advantages for carbonyl relaxation measurements. It allows functionally important carbonyl groups accessible for relaxation measurements *(155)*. Examples of such groups include backbone ^{13}COs of residues preceding proline (Pro); such residues are of biological interest since Pro-rich sequences often comprise the docking sites mediating protein–protein interactions *(156)*. Further examples are the side-chain ^{13}COs of aspartate (Asp) and glutamate (Glu) residues, which participate in electrostatic or hydrogen bond interactions underlying the catalytic mechanisms of therapeutically important enzymes, such as DHFR *(157)*

and HIV–protease *(158)*. All of these ^{13}COs lack an adjacent scalar-coupled ^{15}N–^{1}H moiety, and are therefore beyond the grasp of HNCO-based approaches *(159)*. A second advantage of ^{13}C-detection is that the pulse schemes require fewer extended spin–echo durations, which are typically needed for coherence transfer across small heteronuclear scalar couplings. This means less time spent in the transverse plane, and less signal loss for nuclei plagued by short transverse relaxation times. A third advantage is that peptide linkages that are undetectable via ^{1}H-^{15}N HSQC-based relaxation experiments because of severe exchange ^{1}H and ^{15}N broadening may nevertheless be observable via ^{13}CO-detected methods. Finally, ^{13}C-detection obviously avoids artifacts associated with water suppression and radiation damping. The latter can pose practical problems for ^{1}H-detected relaxation pulse schemes that involve significant proton pulsing to average out cross-correlated relaxation effects and J-coupling artifacts.

In order to address the core research question our first objective will be to determine the efficacy of ^{13}C relaxation studies as a means to detect direct Zn^{2+} coordination. This will be achieved using a simple and inexpensive model complex. It lays the ground to perform relaxation measurements on a series of ^{13}C labeled Protα peptides in the presence and absence of Zn^{2+}.

We have chosen ethylenediaminetetraacetic acid (EDTA) as the model compound for the proof-of-principle ^{13}C relaxation studies (Figure 5.6). EDTA is a well characterized metal chelator with high affinity for Zn^{2+} ($K_d = 10^{-16}$ M) *(160)* and chelates in 1:1 ratio of Zn^{2+}:EDTA.

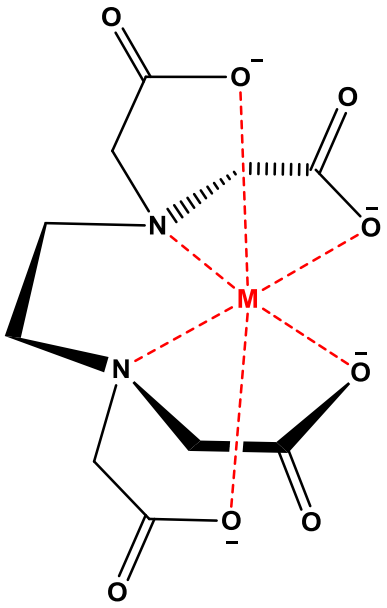


Figure 5.6: EDTA- Zn^{2+} complex showing the ligand coordination and geometry.

The reason for choosing EDTA as a model is because of its strong affinity for Zn^{2+}, structural similarity to the side chain of glutamic acid, and ready availability. The Zn^{2+}-EDTA ion complex is stable this is probably due to the size of the chelate ring. For an assessment of Zn^{2+}-EDTA complexes, several kinds of relaxation data, in the first instance T_1 for EDTA with and without Zn as a function of magnetic field strength and temperature were acquired. From these measurements different contributions to T_1 ($T_{1\ (SR)}$, $T_{1\ (NOE)}$, $T_{1\ (DD)}$, and $T_{1\ (CSA)}$) have been calculated. In order to keep the experimental conditions consistent with Zn^{2+}-Protα studies the spin-lattice (T_1) relaxation times of EDTA with and without Zn^{2+} in 50 mM phosphate buffer, pH 7.4 using 400 MHz NMR were measured. In particular, we focused on the T_1 values for the carbonyl carbons as these are the closest spin-active nuclei to the Zn^{2+} ion and thus should be most sensitive to its presence or absence. However, the relaxation times for every carbon in EDTA were measured.

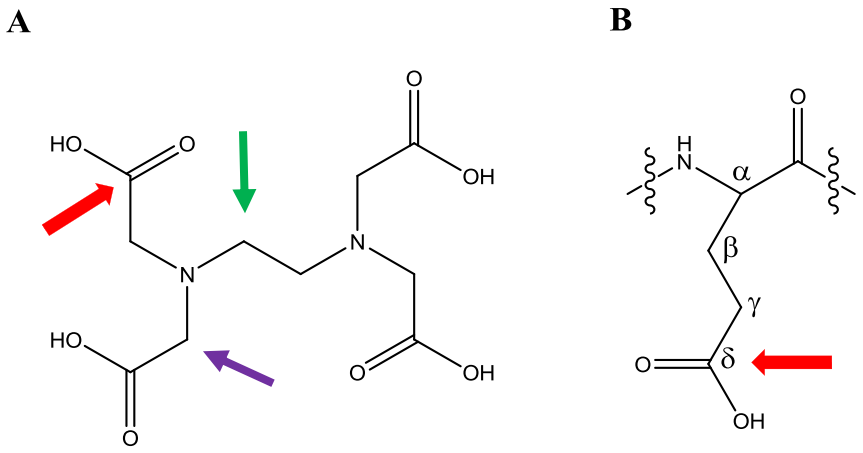


Figure 5.7: A) Structure of EDTA with one of its four carbonyl carbons indicated by red arrow, lateral carbon indicated with a purple arrow, and central carbon indicated with a green arrow. B) Structure of glutamic acid in the context of a peptide sequence. The carbonyl carbon of its carboxyl side chain is indicated with a red arrow.

Expected Results

The T_1 of ^{13}CO group on EDTA should take longer time to relax when there is no Zn^{2+} in the sample. This is because the motion of the CO group is free and rotates faster when there is no Zn^{2+} in the sample. However in the presence of the Zn^{2+} the motion of the CO group is restricted and thereby decreases the relaxation time.

Spin-lattice relaxation measurements were made and the contribution of different relaxation measurements like CSA, DD to the total relaxation were determined. In the case of carbonyl carbon CSA and SR are the dominating mechanisms, as there are no protons to promote the DD mechanism. For lateral and central carbons that have protons attached DD and SR are the dominant relaxations mechanism. Ultimately the correlation times of three carbons with and without Zn^{2+} in the sample were obtained. This study was carried out on 400 and 500 MHz and at three different temperatures (25 ^{o}C, 40 ^{o}C, 50 ^{o}C).

Effect of temperature on correlation times τ_c

The correlation times of the carbons will go down with increase in temperature irrespective of presence or absence of Zn^{2+} in the sample. This is attributed to increase in the molecular motions at high temperatures.

Effect of Zn^{2+} on correlation times

The Zn^{2+} added samples are expected to have longer correlation times irrespective of the temperature because of the restricted motion

Section 5.3 Relaxation rates of the carbonyl carbon

Sample preparation

60 mM of EDTA was dissolved in D_2O to make aqueous solutions of EDTA and the pH was maintained at 7.2. For metalated samples, EDTA - Zn^{2+} aqueous solutions were prepared with 60 mM of EDTA and 90 mM of $ZnCl_2$. All the solutions of EDTA and EDTA- Zn^{2+} were degassed via the freeze thaw method for a minimum of three times to remove any oxygen present in our samples.

^{13}C NMR relaxation measurements were performed at 25 ^{o}C, 40 ^{o}C, 50 ^{o}C. A Bruker 400 MHz (B_0 = 9.4 T) spectrometer at 100.6 MHz and a Varian 500 MHz (B_0 = 11.7 T) spectrometer at 125.2 MHz were used to carry out the measurements. The data was obtained using the inversion recovery method. Non-linear least square procedures were used for the analysis of T_1 measurements. The NOE factors were determined at 125.2 MHz by setting the decoupler on-resonance and far off-resonance in successive experiments.

Theory

EDTA is a simple spin system with three carbon resonances appearing in EDTA at 175.7 ppm (carbonyl), 57.5 ppm (lateral), and 51.2 (central) while in the presence of Zn^{2+} the resonances shifted downfield to 178.1 ppm, 61.1 ppm, and 55.9 ppm respectively. Although there are four carbonyl carbons, two central, and four lateral carbons, only three resonances were observed in the spectrum (in the presence and absence of Zn^{2+}) this indicates they are all chemically equivalent on the NMR time scale. Although the EDTA- Zn^{2+} complex forces the carbons to become chemically inequivalent, this is not observed in the spectrum, indicating that there may be a rapid inter-conversion of the carbonyl groups coordinating Zn^{2+}.

The carbonyl carbon's spin lattice relaxation mechanism is known to relax via the chemical shift anisotropy (R_1^{CSA}) and spin rotation pathways (R_1^{SR}); as given in equation (7):

$$R_1 = R_1^{CSA} + R_1^{SR} \tag{7}$$

Under extreme narrowing conditions and assuming a symmetric shielding tensor, the chemical shift anisotropy is given by equation (8):

$$R_1^{CSA} = \left(\frac{2}{15}\right)\gamma^2 B_O \Delta\sigma^2 \tau_C \tag{8}$$

In eqn. (8), γ is the carbon magnetogyric ratio, B_O is the field strength, $\Delta\sigma$ is the shielding anisotropy (132 ppm in EDTA and 119 ppm in EDTA/Zn^{2+} solution) *(87)* and τ_C is the reorientational correlation time. The spin rotation interaction is given as

$$R_1^{SR} = \left(\frac{8\pi IkT}{h^2}\right)C^2\tau_J \tag{9}$$

where I is the moment of inertial, C is the spin rotation coupling constant, τ_J is the angular momentum correlation time (i.e., time between molecular collisions), k is the Boltzman constant, T is temperature in K, and h is Planck's constant. A full analysis of angular momentum correlation time, through the spin rotation interaction, requires knowledge of the spin rotation coupling constant which is presently unavailable. However, by combining equations (7) and (8), one sees a method by which the SR and CSA contributions can be separated experimentally. This method requires the measurement of the overall relaxation rate on multiple instruments operating at different field strengths; ΔB_O.

$$R_1 = \left(\frac{2}{15}\right)\gamma^2 \Delta B_O \Delta\sigma^2 \tau_C + R_1^{SR} \quad (10)$$

A linear fit of R_1 versus ΔB_O will yield the reorientational correlation time from the slope. The SR contribution can be deduced from either the intercept and/or from the difference between R_1 and R_1^{CSA}.

Calculation of Activation energies:

Activation energy of the three carbons in EDTA is calculated before and after Zn^{2+} addition using the Arrhenius equation.

$$k = A.e^{-E_{a/RT}} \quad (11)$$

k is the rate constant, A is pre-exponential factor, E_a is the activation energy, and R is the universal gas constant. Taking the logarithm on both sides yields the equation above in the form of the equation of a straight line (y = mx+b):

$$lnk = -\frac{E_a}{RT} + lnA \quad (12)$$

and all the variables can be found: y = lnk, m(slope) = $-E_a/R$, x = 1/T, and b = lnA

An Arrhenius plot of τ_C (y-variable) versus T^{-1} (x-variable) gives a negative slope from which activation energies will be obtained using the equation

$$E_a = -slope\ of\ arrhenius\ plot * R \quad (13)$$

Where R is the gas constant (8.3 J/mol*K)

These experiments will reveal if the changes in relaxation time are large enough for us to measure and thus tell if Zn^{2+} binding is taking place and will stand as a strong proof of principle to study Zn^{2+} binding in Protα.

Results

The various relaxation contributions of the carbonyl carbons along with the reorientational times are given in Tables 5.1 and 5.2.

Table 5.1 shows the variable temperature relaxation rates (R_1), the chemical shift anisotropy (R_1^{CSA}) and spin rotation (R_1^{SR}) contributions as well as the correlation times (τ_C) calculated via equation (7) of the carbonyl carbon of EDTA in the absence of Zn^{2+}. As expected, these rates are seen to decrease with rising temperature indicating a decrease in the effectiveness of the dominating mechanism; CSA. The SR mechanism displays the opposite trend; increasing with rising temperature. The last column of this table contains the values for the correlation times obtained via the CSA contribution and equation (8). The correlation times, τ_C, are seen to decrease with rising temperature indicating faster molecular dynamics at the carbonyl site with increasing temperature.

Table 5.2 lists the same variable temperature data but with the addition of Zn^{2+} to the solution. A comparison of the relaxation rates from these two tables reveals that the introduction of Zn^{2+} to the solution does affect the relaxation rates of the carbonyl carbon appreciably. One notices an increase in the overall relaxation rate and a similar tendency for the CSA contribution. At the lower field strength, the data show the SR mechanism dominates at 323K. A comparison of the correlation times in Tables 5.1 and 5.2 indicate that the presence of Zn^{2+} does lead to noticeable reduction of the molecular dynamics at the carbonyl site indicating intermolecular interactions at

this site with Zn^{2+}. This observation is further supported by the differences in the activation energies at this site. An Arrhenius fit of τ_c versus T^{-1} yielded energies of activation of 33.7 and 40.3 kJ/mole without and with Zn^{2+}, respectively; suggesting the 6.7 kJ/mol energy difference is due to the presence of intermolecular interactions between Zn^{2+} and the carbonyl sites.

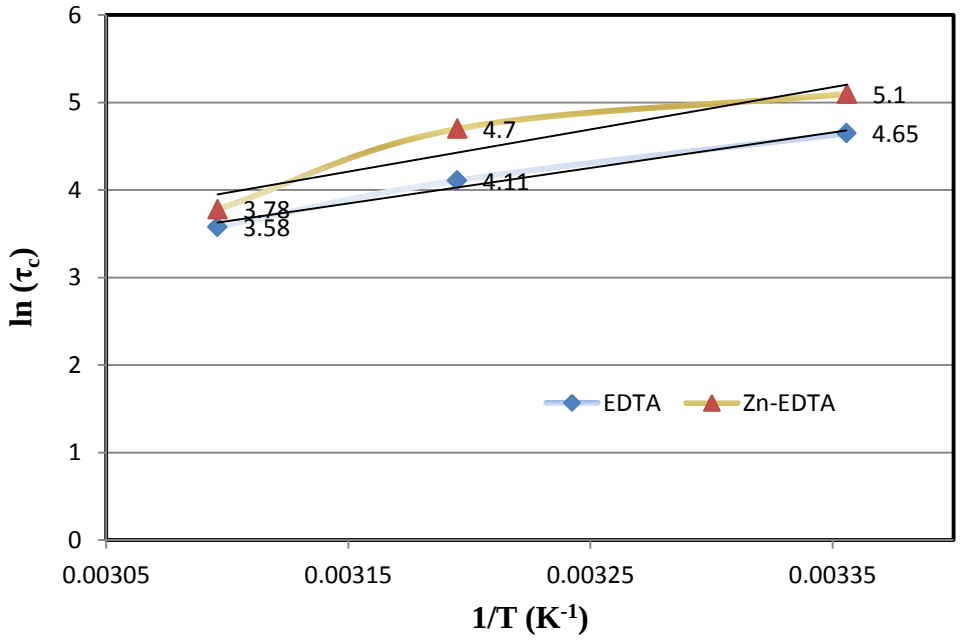


Figure 5.8: Arrhenius plot of carbonyl carbon in EDTA with and without Zn^{2+}. The slopes were used to calculate activation energies.

Table 5.1 Spin Lattice relaxation rates, CSA and SR contributions, and correlation times of the carbonyl carbon of EDTA at different field strengths and as a function of temperature.

T(K)	B_o(MHz)	R_1(1/s)	$R_1^{(CSA)}$(1/s)	$R_1^{(SR)}$ (1/s)	%CSA	%SR	τ_C (ps)
298	125.2	0.175	0.150	0.025	85.7	14.3	105
	100.6	0.104	0.082	0.022	78.8	21.2	
313	125.2	0.113	0.088	0.025	77.9	22.1	61
	100.6	0.081	0.056	0.025	69.1	30.9	
323	125.2	0.087	0.053	0.034	60.9	39.1	36
	100.6	0.068	0.034	0.034	50.0	50.0	

Table 5.2: Spin Lattice relaxation rates, CSA and SR contributions, and correlation times of the carbonyl carbon of EDTA with Zn^{2+} at different field strengths and as a function of temperature

T(K)	B_o(MHz)	R_1(1/s)	$R_1^{(CSA)}$ (1/s)	$R_1^{(SR)}$ (1/s)	%CSA	%SR	τ_C (ps)
298	125.2	0.214	0.195	0.019	91.1	8.9	165
	100.6	0.130	0.111	0.019	85.4	14.6	
313	125.2	0.135	0.129	0.006	95.6	4.4	110
	100.6	0.089	0.083	0.006	93.3	6.7	
323	125.2	0.094	0.051	0.043	54.3	45.7	44
	100.6	0.076	0.033	0.043	43.4	56.6	

Section 5.4 Relaxation rates of the lateral and central carbons

Theory

The lateral and central carbons possess two hydrogens each which will lead to dipole-dipole interactions. The spin lattice relaxation rate of these two carbons proceeds primarily via the dipole-dipole (R_1^{DD}) and spin rotation (R_1^{SR}) mechanisms.

$$R_1 = R_1^{DD} + R_1^{SR} \tag{14}$$

The dipole-dipole contribution can be experimentally isolated through the use of NOE measurements and equation (15):

$$R_1^{DD} = \frac{\eta_{NOE}}{1.987} R_1 \tag{15}$$

The spin rotation contribution is given by equation (5) but can be obtained from the difference between R_1 and R_1^{DD}. The R_1^{DD} and R_1^{SR} relaxation contributions for the lateral and central carbons, along with the re-orientational correlation times, are given in Tables 5.3 and 5.4.

Results and discussion

Table 5.3 lists the overall spin lattice relaxation rates (R_1) of the lateral and central carbons, the NOE values (η_{NOE}), the dipole-dipole (R_1^{DD}) and spin rotation (R_1^{SR}) contributions as well as the correlation times (τ_c) at these two sites. This table illustrates the overall relaxation rates decrease only slightly with rising temperature with the dipole-dipole mechanism dominating at all temperatures. As anticipated, the spin rotation contribution gradually increases with rising temperature. The final column in this table lists the correlation times which provide a glimpse of

the dynamics at these two sites. One sees a larger increase in the reorientational motion of the lateral carbon with rising temperature as compared to the central carbon. An Arrhenius fit of the correlation times versus T^{-1} yielded energies of activation of 22.1 and 9.7 kJ/mole for the lateral and central carbons, respectively. These activation energies indicate greater rotational freedom at the central versus the lateral carbon site.

The experimental spin lattice relaxation rates (R_1), the NOE values (η_{NOE}), the dipole-dipole (R_1^{DD}) and spin rotation (R_1^{SR}) contributions as well as the correlation times (τ_c) of the lateral and central carbons upon the introduction of Zn^{2+} to the EDTA solution are found in Table 5.4. The data in this table illustrate a more dramatic change in all the experimental observables with the introduction of Zn^{2+} to the solution. In particular, the overall relaxation rates of the lateral carbon experience larger changes with increasing temperature. In fact, and surprisingly, the dipole-dipole mechanism becomes less efficient than spin rotation at all temperatures for this lateral carbon suggesting an increase in the re-orientational motion at this carbon site which is supported by the reduction of τ_c values with the introduction of Zn^{2+}. This observation is conceivable if EDTA and Zn^{2+} interactions are simultaneously occurring at the carbonyl and $-CH_2$-N- (central site) resulting in enhanced rotational freedom (i.e. "quasi" free spinning) at the lateral methylene carbon site. An Arrhenius fit of this carbon's correlation times versus T^{-1} yielded an energy of activation of 20.4 kJ/mole; approximately 2.0 kJ/mol less than without the presence of Zn^{2+}.

The overall relaxation rate of the central methylene group decreases incrementally with rising temperature with the dipole-dipole mechanism dominating at all temperatures. A comparison of the correlation times, without (Table 5.3) and with Zn^{2+} (Table 5.4), shows a slowing of the

reorientational motion of this methylene group upon the introduction of the Zn^{2+} ion. The τ_c data indicate the presence of intermolecular interactions between Zn^{2+} and these $-CH_2$-N- sites. The Arrhenius fit of the central carbon's correlation times versus T^{-1} yield activation energy of 11.4 kJ/mole. One sees an increase of 1.7 kJ/mol in the activation energy upon the introduction Zn^{2+} to the solution indicating the presence of intermolecular interactions at the $-CH_2$-N- sites.

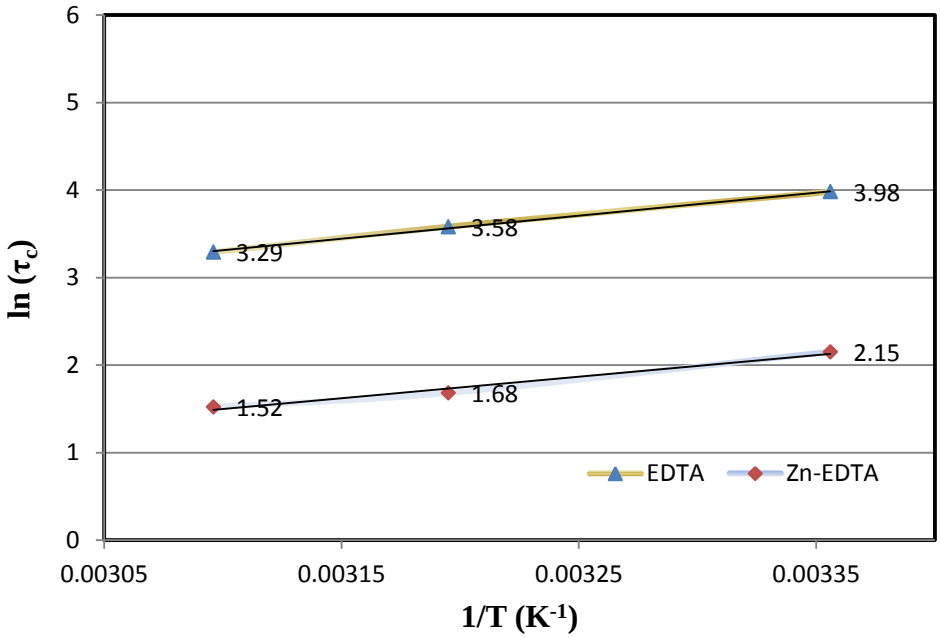


Figure 5.9: Arrhenius plot of lateral carbon in EDTA with and without Zn^{2+}. The slopes were used to calculate activation energies.

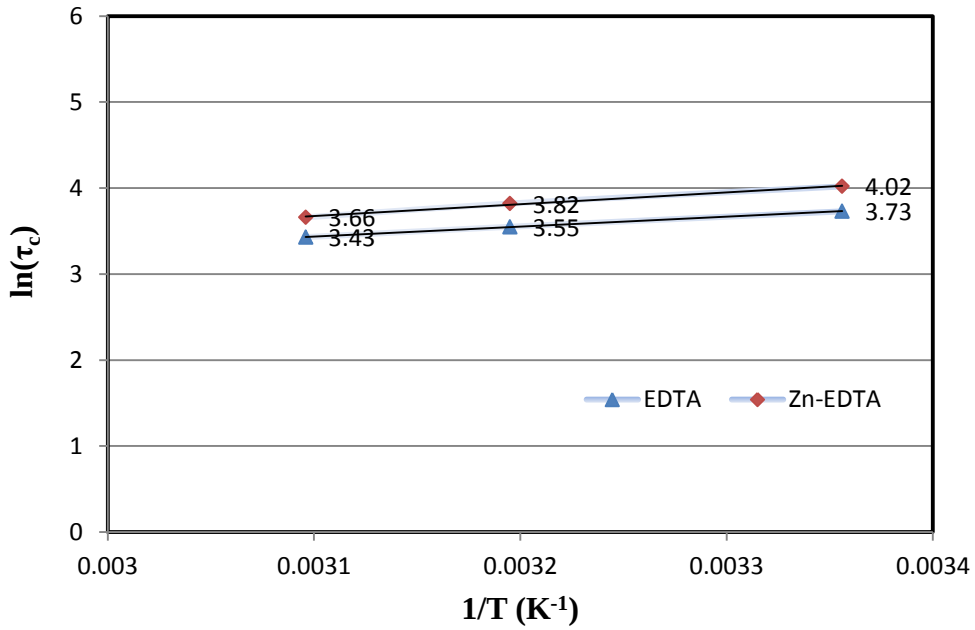


Figure 5.10: Arrhenius plot of central carbon in EDTA with and without Zn^{2+} showing the slope. The slopes were used to calculate activation energies.

Table 5.3 Spin Lattice relaxation rates, NOE, dipole-dipole and spin rotation contributions, and correlation times of the lateral and central carbons of EDTA as a function of temperature at 125.2 MHz (11.7 T).

T(K)	carbon	R_1(1/s)	η_{NOE}	R_1^{DD}(1/s)	R_1^{SR}(1/s)	%D	%SR	τ_C (ps)
298	Lateral	2.94	1.57	2.32	0.62	78.9	21.1	54
	Central	2.94	1.22	1.80	1.14	61.2	38.8	42
313	Lateral	2.04	1.51	1.55	0.49	76.0	24.0	36
	Central	2.04	1.49	1.53	0.51	75.0	25.0	35
323	Lateral	1.61	1.44	1.17	0.44	72.7	27.3	27
	Central	2.00	1.33	1.34	0.66	67.0	33.0	31

Table 5.4: Spin Lattice relaxation rates, NOE, dipole-dipole and spin rotation contributions, and correlation times of lateral and central carbons of EDTA with Zn^{2+} as a function of temperature at 125.2 MHz (11.7 T).

T(K)	carbon	R_1(1/s)	η_{NOE}	$R_1^{(DD)}$(1/s)	$R_1^{(SR)}$ (1/s)	%DD	%SR	τ_C (ps)
298	Lateral	3.13	0.24	0.37	2.76	11.8	88.2	8.6
	Central	4.00	1.19	2.40	1.60	60.0	40.0	56
313	Lateral	1.30	0.35	0.23	1.07	17.7	82.3	5.4
	Central	3.23	1.21	1.97	1.26	61.0	39.0	46
323	Lateral	0.73	0.54	0.20	0.53	27.4	72.6	4.6

Section 5.5 Conclusions

The study was undertaken to explore the effectiveness of spin inversion recovery experiments to detect Zn binding in Protα. EDTA and Zn complexes were originally chosen because of the structural similarity of EDTA to the carboxylate group of aspartic and glutamic acid, which are the plausible ligands of Zn in Protα(50-89)N50W. The general conclusion from our work is that a difference in the relaxation times was successfully detected in the EDTA samples with and without Zn, thus providing a proof of principle to carry out Zn binding study on Protα(50-89)N50W or other acidic sequences. It should be noted that in order to get residue level information peptides will require ^{13}C labeling.

To get into more details about the work, spin lattice relaxation measurements conducted on EDTA/D_2O solutions as well as EDTA/Zn^{2+}/D_2O solutions as a function of field strength and temperature were used to calculate the contribution of different relaxation methods (CSA, SR, DD) and correlation times. These measurements showed that the chemical shift anisotropy mechanism is the dominating mechanism in the carbonyl carbon relaxation process. A comparison of the correlation times in solutions containing Zn^{2+} and without Zn^{2+} showed the presence of intermolecular interactions at these sites. Surprisingly, our results show that the lateral methylene groups increase their reorientational motion upon the introduction of Zn^{2+} to the solution suggesting the anchoring of the carbonyl and the central methylene sites give rise to increased reorientational freedom at the lateral methylene sites.

The results obtained here coincide well with a previous study of EDTA and Zn carried out at pH 12.0 *(87).* In both the studies there was an increase in the correlation times of central and

carbonyl carbon before and after Zn addition. However, the lateral carbon showed a decrease in correlation times in the presence of Zn (compared to the no Zn sample) which deviates from the previous study. This could be due to the difference in the pH in the two studies. At pH 12.0 all the carboxyl groups remain deprotonated and can bind Zn more strongly whereas at pH 7.4 the carboxyl groups of EDTA may only be partially deprotonated potentially allowing for exchange of the carboxylate groups between the axial and equatorial positions causing the difference in the correlation times. This study tells us the depth of information that can be obtained for Zn binding in Protα(50-89)N50W.

Chapter 6 Conclusions and Future Studies

The central research question presented here is to understand if natively unfolded proteins bind metals like folded proteins that being with a well-defined set of ligands, or if there is more heterogeneity in ligand set used. Protα is a natively unfolded protein that binds metal cations but undergoes structural rearrangement only in the presence of Zn^{2+}. Zn^{2+} binding in Protα was taken as a case study to understand metal binding in natively unfolded proteins.

The following research questions have been addressed in our research to contribute in the investigation of Zn^{2+} binding region of Protα.

1.1 Develop a synthetic method using 4-pyridinemethanol as a side chain protecting group on Asp and Glu to improve the synthesis, purification, and characterization of highly negatively charged sequences.

1.2 Determine the structural changes in the central segment of Protα in the presence of environmental factors like Zn^{2+} and temperature.

1.3 Investigate the structural changes in the Protα(50-89)N50W using differential scanning calorimetry (DSC).

1.4 Further refine the map of the Zinc binding region of Protα.

The results obtained for each of these questions have been summarized below.

The 4-pyridinemethanol protected aspartic and glutamic acid were developed to aid in the synthesis, purification, and characterization of central metal binding 50-89 region of Protα. Use of this protecting group in conjunction with benzyl ester side chain protection allows for acidic peptides to be separated using RP HPLC and characterized by positive ion mode ESI-MS. CD

studies have shown that majority of the Zn^{2+} binding occurs in the 50-89 region and that the major structural changes occur in this region. CD data has also shown that the peptide attains a secondary structure upon heating from 20–80 ^{o}C. DSC data also showed that the peptide undergoes an exothermic event upon heating suggesting an increase in structure which supports the temperature data obtained from CD. NMR studies on EDTA and Zn^{2+} studies have shown the effectiveness of ^{13}C relaxation as a means to detect direct metal coordination.

Significance:

Difficult syntheses, such as that of ProTα(48-110), may now be feasible by a linear synthesis with use of an OMePyr protecting strategy. This allows researchers to study a larger variety of potential systems and can allow one to synthesize isotopically labeled peptides. It is important to have very robust synthetic methods to make these as isotopes are expensive. The fact that a highly charged sequence like ProTα(50-89) can selectively bind Zn^{2+} and becomes ordered when it does so provides insight into general metalloprotein design principles. Specifically, this can be used for selective disorder-to-order transitions which may serve a biological purpose such as enhancing binding to a partner (e.g. Rev or CREB). Thus, metal binding can be used as an on/off signal for proteins like ProTα.

Limitations:

1. The 4-pyridinemethanol protected aspartic and glutamic acid were not applied in the synthesis of Protα(50-89)N50W segment. Application of the group in model peptide has shown how the new protecting group performs in peptide synthesis protocols. However, long synthesis, purification, and characterization of long acidic sequences has always

been challenging and there is still a need to look at the performance of the new protecting group in the synthesis, purification and characterization of long peptide sequences.

2. Although the data obtained from the secondary structural changes in Protα(50-89)N50W studied on CD and DSC by adding Zn^{2+} and varying temperature is informative, additional experiments that help us obtain more thermodynamic information are required to understand the details of temperature induced structural changes and to get a deeper insight into why the thermogram of Zn^{2+} added peptide showed a broader exothermic event compared to the no Zn^{2+} sample.

To overcome these limitations the future work on this project should be focused on synthesizing the 50-89 sequence of Protα using 4-pyridinemethanol protected aspartic and glutamic acid. In terms of building a thermodynamic profile of temperature induced structural changes in the peptide with and without Zn^{2+} added, an NMR study measuring the spin-lattice relaxation times (T_1) would be helpful. From the relaxation measurements and ^{13}C labeling studies one could also address the question if Zn^{2+} binding in Protα occurs at a specific site or if several degenerate sites exist.

Relaxation measurements with selective ^{13}C isotope labeling of carboxyl groups in aspartic and glutamic acid is the next step to investigate Zn^{2+} binding in Protα. Aspartic and glutamic acid will be selectively labeled with ^{13}C isotope and a series of ^{13}C NMR spectra will be collected using the inversion recovery method. Through creation of a series of selectively ^{13}C labeled peptides, these studies will show if Zn^{2+} binding occurs to specific set of residues or if several degenerate sites exist. The existence of degenerate metal binding sites in at least two other IDPs has been noted *(163, 164)* and if this finding bears out in Protα it would suggest this may be a general feature of metallo-IDPs.

Chapter 7 Materials and Methods

Materials for peptide synthesis

Rink amide MBHA resin, HBTU, HOBt and Fmoc-amino acids were purchased from EMD Millipore. *N*,*N*-Diisopropylethylamine (DIEA), triisopropylsilane (TIS), 4-pyridinemethanol, and Pd/C were obtained from Sigma-Aldrich. Both *N*,*N*-dimethylformamide (DMF) and trifluoroacetic acid (TFA) were purchased from Fisher Scientific.

Methods for model peptide synthesis using 4-pyridylmethyl esters *(161)*

Peptides were synthesized using a Protein Technologies Inc. PS3 peptide synthesizer. All the peptides were synthesized on the 0.1 mmol scale using Rink amide MBHA resin with 0.3 mmol/g substitution. Upon cleavage, the peptide produced from this resin was *C*-terminal amidated. Coupling of each amino acid was carried out in DMF using 4 equiv Fmoc-amino acid/HBTU/HOBt and 4 equiv DIEA with a 1 h coupling time. When compounds **3a** or **3b** were incorporated into a peptide 2 equiv of Fmoc-amino acid/HBTU/HOBt were used. The Fmoc group was removed using 20% v/v piperidine in DMF with a reaction time of 30 min. The final peptide was acetylated using 2 mL of acetic anhydride and 4 equiv DIEA in DMF.

After synthesis the resin was washed with dichloromethane, methanol and then dried. The resin was dried for 2 minutes using vacuum. The dried resin was treated with a TFA/TIS/H2O (95/2.5/2.5) solution for 2 h. After 2 hours the reaction mixture was filtered and the peptide precipitated with cold diethyl ether. The precipitate was collected, dissolved in water and freeze dried. The peptide was purified and characterized by RPHPLC and ESI-MS respectively and mass of the dried peptide was obtained to calculate percentage yields.

Purification by HPLC

Peptides were purified on a Vydac C18 column (5 μm, 4.6 x 250 mm) using a BioRad HPLC system. Crude peptides were purified using 0.1% TFA in water (Buffer A) and 0.1% TFA in acetonitrile (Buffer B). After injection, isocratic elution at 100% Buffer A was run for 3 min. followed by gradient elution from 100% Buffer A to 0% Buffer A in 37 min. with a flow rate of 1.0 mL/min. Peptide signals were detected at both 214 nm and 280 nm.

Characterization

All peptides were characterized by ESI-MS after elution from HPLC, and thus were dissolved in water/acetonitrile containing 0.1% TFA. Sample was delivered to the ESI source with a KD Scientific syringe pump at 10 μL/min, and the electric potential used to initiate ESI (i.e. capillary voltage) was 2500 V. The source temperature was set at 90 °C and the desolvation temperature at 275 ^{o}C with a cone gas flow of 10 L/h and a desolvation gas flow of 500 L/h.

Hydrogenation

The peptide was dissolved in 10 mL of ethyl acetate and 2 mg of 5% Pd/C was added. The reaction flask was placed under a hydrogen atmosphere of 70 psi and shaken overnight using a Parr hydrogenator. The Pd/C catalyst was removed by filtration through a fritted glass filter and the ethyl acetate was removed under reduced pressure.

Materials for synthesis of Protα(50-89)N50W and Polyglutamic acid

Rink Amide MBHA resin, HBTU, HOBt, and Fmoc-amino acids were purchased from NovaBioChem. N,N-diisopropylethylamine (DIEA) and triisopropylsilane were obtained from Aldrich. N,N-dimethylfomamide and trifluoroacetic acid were purchased from Fischer scientific. Protα(50-89)N50W and Polyglutamic acid were purified on a mono-Q 5/5 column using a

BioRad HPLC system. Slide-A-Lyzer cassettes with 3.5 kD MWCO with 0.5-3.0 mL and 4.0-12.0 mL volumes were obtained from Pierce Biotechnology.

Methods for synthesis of Protα(50-89)N50W and Polyglutamic acid

Synthesis of Protα(50-89)N50W was performed manually using a CEM Discover Microwave Synthesizer. All the amino acids were assembled in a linear fashion on a 0.25 mmol scale using Rink Amide MBHA resin with 0.35 mmol/g substitution. All the Fmoc amino acids used in the synthesis were Fmoc protected at the N-terminal end and side chain protected where applicable. The Fmoc Protecting group was removed using 20% (v/v) solution of piperidine in DMF and the reaction time was for 3 minutes. Coupling reactions were performed in DMF by *in situ* activation of a 4 equivalent excess of Fmoc-amino acid using 8 equivalents excess of HBTU/HOBt/DIEA. Coupling times were for 5 minutes for all the amino acids. After each coupling and deprotection step the resin was extensively rinsed with DMF and the completeness of these reactions was often tested using the ninhydrin test. The final peptide sequence was acetylated using acetic anhydride in DMF with 2.0 mmol DIEA.

Synthesis of the Polyglutamic acid sequence was carried out on a PS3 synthesizer. All the amino acids were arranged in a linear fashion on 0.1 mmol scale Rink Amide MBHA resin with 0.35 mmol/g substitution. All the solvents used in the synthesis were the same as those used in microwave synthesis mentioned above. The coupling and deprotection times were usually 1.5 hours and 30 min respectively for each amino acid unless the reaction was deemed incomplete. The Fmoc protecting group was removed from the amino acid by using 20% piperidine. The final peptide was acetylated using acetic anhydride in DMF with 2.0 mmol DIEA.

The glutamic acid residues used in the synthesis of Protα(50-89)N50W and Polyglutamic acid were tertiary butyl (tBu) protected.

Upon on completion of synthesis of the target sequences, the resins from individual synthesis were washed with DMF, dichloromethane and methanol separately. They were dried for 2 minutes using vacuum. Each peptide was cleaved from the resin in a solution of TIS/H_2O/TFA (2.5/2.5/9.5) for 2 hours. After 2 hours the reaction mixture was collected and peptide was precipitated with cold diethyl ether. The precipitate was obtained as a pellet after centrifugation and decanting the supernatant. The crude peptide was dissolved in water, flash frozen and freeze dried.

Purification by HPLC

Protα(50-89)N50W and Polyglutamic acid were purified on a mono-Q 5/5 strong anion exchange (SAX) column using 20 mM glycine with 1M NaCl at pH 5.3 as buffer A and 20 mM glycine at pH 5.3 as buffer B. A gradient elution from 100% B to 0% B in 32 minutes was used and Protα(50-89)N50W and polyglutamic acid had a retention time of 18 minutes.

Dialysis

The peaks collected from SAX were dialyzed extensively to get rid of any sodium in the samples. The peptide samples were loaded into 2 kD MWCO dialysis cassettes (0.5 – 3.0 mL or 4.0 – 12.0 mL capacities) and dialyzed against 1 L of 150 mM ammonium acetate three times for 1 hour each, later dialyzed against 50 mM ammonium formate thrice for 1 hour each and against 1 L of deionized water (18 MΩ cm^{-1}, pH = 5.1) once for 2 hours.

Characterization

Samples from Protα(50-89)N50W and Polyglutamic acid sequence were characterized on a Micromass Q-ToF Micro Mass Spectrometer in negative ion mode. Purified and dialyzed peptides were delivered to the ESI source with a KD scientific syringe pump at 10 μl/min, and the electric potential used to initiate ESI was 2600 V. The source conditions were optimized to provide best S/N ratio with lowest amount of fragmentation (facile losses of water). The source temperature was set at 50 ^{o}C. The nebulizer gas pressure was 25 psi and the dry gas flow rate was 10 mL/min. The capillary skimmer voltage difference was 88.5 V.

CD Experiments

50 mM Phosphate buffer

Sodium biphosphate and sodium monophosphate were purchased from Sigma Aldrich. 18 mΩ-cm water was used to make the buffer. The pH was adjusted to 7.2.

Zn^{2+} studies

CD spectra were obtained by acquiring 8 scans from 260 to 190 nm on a Jasco J-815 spectrometer (equipped with Jasco SOFTWARE V) using a 1 mm pathlength quartz cell (Starna Cells, Atascadero, CA) at a speed of 20 nm/min, a 1 nm bandwidth, a data pitch of 0.2 nm, response time 0.25 sec and 5 mdeg sensitivity at room temperature (~25 ºC) in a continuous scanning mode. Contributions due to phosphate buffer were eliminated by subtracting the phosphate buffer spectrum from the peptide solutions. All analysis of CD spectra were conducted after smoothing (Means-movement with convolution width of 5-25) and conversion to molar ellipticity using the JASCO Spectra Analysis program.

Peptide Concentration (mM)	$ZnCl_2$ Concentration (mM)	Eq. of Zn^{2+} in peptide sample
0.5	0	0
0.5	10	20
0.5	17.5	35
0.5	25	50
0.5	37.5	75
0.5	50	100
0.5	75	150
0.5	100	200

Table 7.1: Concentrations of the peptide and concentration and equivalents of Zn^{2+} used in Zn^{2+} titrated CD runs

Temperature studies

CD spectra were obtained by acquiring 8 scans from 260 to 190 nm on a Jasco J-815 spectrometer (equipped with Jasco SOFTWARE V) equipped with temperature cell holder using a 0.1 mm pathlength cylindrical quartz cell (Starna Cells, Atascadero, CA) at a speed of 20 nm/min, a 1 nm bandwidth, a data pitch of 0.2 nm, response time 0.25 sec and 5 mdeg sensitivity in a temperature range of 20-80 ^{o}C in a continuous scanning mode. Contributions due to phosphate buffer were eliminated by subtracting the phosphate buffer spectrum from the

peptide solutions at respective temperatures. All analysis of CD spectra were conducted after smoothing (Means-movement with convolution width of 5-25) and conversion to molar ellipticity using the JASCO Spectra Analysis program.

Modulated DSC Experiments

Sample preparation for DSC studies

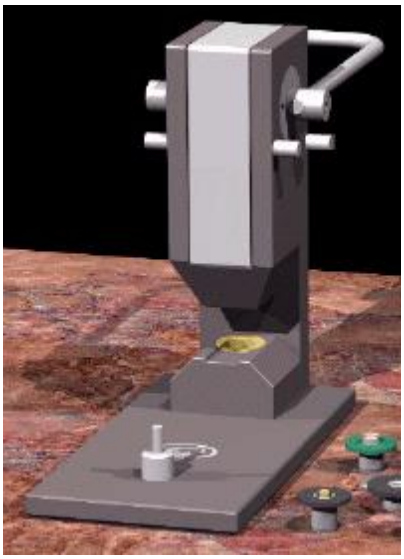


Figure 7.1: T zero sample press and pans used to seal samples for DSC experiments.

DSC experiments were carried out on TA Instruments Q2000 calorimeter. 50 mM phosphate buffer at pH 7.2 was used as a solvent to prepare peptide solutions. Prior to analysis the samples were centrifuged for 5 min. Protein concentration was maintained at 50 mg/ml. Thermograms were collected on a 10 µl sample of Protα(50-89)N50W with and without Zn^{2+} in a hermetically sealed t-zero aluminium pans, at a scan rate of 1^{o}C/min from 25 to 80 ^{o}C under modulated conditions. Data was analyzed using universal analysis (TA instruments), to determine the melting temperature (T_m) and enthalpy (ΔH) associated with the thermal events in Protα peptide in the presence of Zn^{2+}, using a horizontal baseline.

EDTA and Zn^{2+} studies

Materials

Ethylenediaminetetraacetic acid disodium salt dehydrate, D_2O, and $ZnCl_2$ were purchased from Sigma aldrich (>99%).

Methods

60 mM of the salt was dissolved in D_2O to make aqueous solutions of EDTA. EDTA- Zn^{2+} aqueous solution was prepared with 60 mM of EDTA and 90 mM of $ZnCl_2$. The pH of both the solutions was maintained at 7.2. All the solutions of EDTA and EDTA- Zn^{2+} were degassed via the freeze thaw method for a minimum of three times to remove any oxygen present in our samples.

NMR measurements

^{13}C NMR relaxation measurements were performed at 25 ^0C, 40 ^0C, 50 ^0C. Bruker 400 MHz (B_0=9.4 T) spectrometer at 100.6 MHz and Varian 500 MHz (B_0 =11.7 T) spectrometer at 125.2 MHz were used to carry out the measurements. The data was obtained using Inversion recovery method. Non-linear least square procedures were used for the analysis of T_1 measurements. The NOE factors were determined at 125.2 MHz by setting the decoupler on-resonance and far off-resonance in successive experiments.

Freeze thaw procedure

The sample was placed in the NMR tube shown above to which a ground glass joint had been attached. The tube is then attached to a vacuum manifold and the sample was frozen using liquid nitrogen.

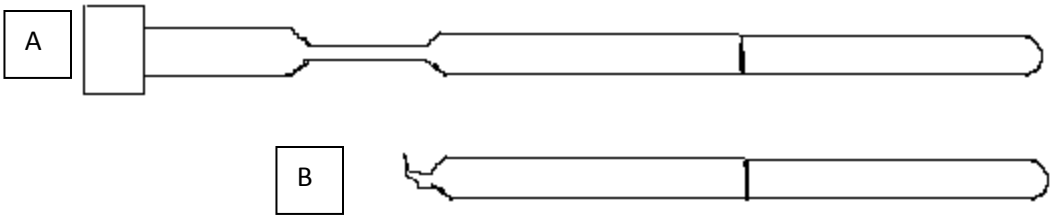


Figure 7.2: NMR sample tubes used to seal the sample by freeze thaw procedure. A) Sample tube before sealing B) sample tube after sealing.

After freezing the sample it was pumped on for 1 min and allowed to thaw slowly under its own vapor pressure. The procedure was repeated at least 2 more times. The process is tricky because if the liquid nitrogen touches the NMR tube it might break. So the sample should be frozen using liquid nitrogen yet not bringing into contact with it.

Chapter 8

Bibliography

1. Fink, A. L. Natively unfolded proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 35-41.

2. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001**, *19*, 26-59.

3. Pullen, R. A.; Jenkins, J. A.; Tickle, I. J.; Wood, S. P.; Blundell, T. L. The relation of polypeptide hormone structure and flexibility to receptor binding: the relevance of X-ray studies on insulins, glucagon and human placental lactogen. *Mol. Cell. Biochem.* **1975**, *8*, 5-20.

4. Cary, P. D.; Moss, T.; Bradbury, E. M. High-resolution proton-magnetic-resonance studies of chromatin core particles. *Eur. J. Biochem.* **1978**, *89*, 475-482.

5. Schweers, O.; Schonbrunn-Hanebeck, E.; Marx, A.; Mandelkow, E. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *J. Biol. Chem.* **1994**, *269*, 24290-24297.

6. Weinreb, P. H.; Zhen, W.; Poon, A. W.; Conway, K. A.; Lansbury, P. T.,Jr NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **1996**, *35*, 13709-13715.

7. Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321-331.

8. - Holt, C.; - Sawyer, L. - Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the [small alpha]S1-, [small beta]- and [small kappa]-caseins. *J. Chem. Soc., Faraday Trans.* , **1993**; 89(15): 2683-92.

9. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK. Natively Disordered Proteins. In: JK Buchner T, editor. Protein Folding Handbook. Vol. I. KGaA, Weinheim, Germany: *Wiley-VCH Verlag GmbH & Co; **2005***. pp. 275–357.

10. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K. Predicting intrinsic disorder from amino acid sequence. *Proteins* **2003**, *53 Suppl 6*, 566-572.

11. Uversky, V. N. What does it mean to be natively unfolded? *Eur. J. Biochem.* **2002**, *269*, 2-12.

12. Uversky, V. N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **2002**, *11*, 739-756.

13. Uversky, V. N.; Radivojac, P.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K. Prediction of intrinsic disorder and its use in functional proteomics. *Methods Mol. Biol.* **2007**, *408*, 69-92.

14. Romero, P.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E.; Dunker, A. K. In *In Identifying disordered regions in proteins from amino acid sequence;* Neural Networks,1997, International Conference; 1997; Vol. 1, pp 90-95 vol. 1.

15. Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Sequence complexity of disordered protein. *Proteins* **2001**, *42*, 38-48.

16. Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415-427.

17. Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* **2000**, *11*, 161-171.

18. Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K. Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **2005**, *44*, 1989-2000.

19. Kozlowski, L. P.; Bujnicki, J. M. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **2012**, *13*, 111-2105-13-111.

20. Lin, ,Chin-Teng; Lin, ,Ken-Li; Yang, ,Chih-Hsien; Chung, ,I.-Fang; Huang, ,Chuen-Der; Yang, ,Yuh-Shyong Protein Metal Binding Residue Prediction Based on Neural Networks. *Int. J. Neur. Syst.* **(2005)**, 15, 71-77.

21. Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23-30.

22. Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J. R.; Jensen, M. R.; Segard, S.; Bernado, P.; Charavay, C.; Blackledge, M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **2012**, *28*, 1463-1470.

23. Fisher, C. K.; Stultz, C. M. Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2011**, *21*, 426-431.

24. Schneider, R.; Huang, J. R.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon, L.; Jensen, M. R.; Blackledge, M. Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. Biosyst.* **2012**, *8*, 58-68.

25. Rezaei-Ghaleh, N.; Blackledge, M.; Zweckstetter, M. Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *Chembiochem.* **2012**, *13*, 930-950.

26. Dyson, H. J.; Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **2004**, *104*, 3607-3622.

27. Tompa, P. On the supertertiary structure of proteins. *Nat. Chem. Biol.* **2012**, *8*, 597-600.

28. McCann, J. J.; Zheng, L.; Rohrbeck, D.; Felekyan, S.; Kuhnemuth, R.; Sutton, R. B.; Seidel, C. A.; Bowen, M. E. Supertertiary structure of the synaptic MAGuK scaffold proteins is conserved. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15775-15780.

29. Zhang, J.; Lewis, S. M.; Kuhlman, B.; Lee, A. L. Supertertiary structure of the MAGUK core from PSD-95. *Structure* **2013**, *21*, 402-413.

30. Uversky, V. N.; Li, J.; Fink, A. L. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* **2001**, *276*, 10737-10744.

31. Bhattacharyya, J.; Das, K. P. Molecular chaperone-like properties of an unfolded protein, alpha(s)-casein. *J. Biol. Chem.* **1999**, *274*, 15505-15509.

32. Thomas, J.; Van Patten, S. M.; Howard, P.; Day, K. H.; Mitchell, R. D.; Sosnick, T.; Trewhella, J.; Walsh, D. A.; Maurer, R. A. Expression in Escherichia coli and characterization of the heat-stable inhibitor of the cAMP-dependent protein kinase. *J. Biol. Chem.* **1991**, *266*, 10906-10911.

33. Uversky, V. N.; Karnoup, A. S.; Segel, D. J.; Seshadri, S.; Doniach, S.; Fink, A. L. Anion-induced folding of Staphylococcal nuclease: characterization of multiple equilibrium partially folded intermediates. *J. Mol. Biol.* **1998**, *278*, 879-894.

34. Semisotnov, G. V.; Kihara, H.; Kotova, N. V.; Kimura, K.; Amemiya, Y.; Wakabayashi, K.; Serdyuk, I. N.; Timchenko, A. A.; Chiba, K.; Nikaido, K.; Ikura, T.; Kuwajima, K. Protein globularization during folding. A study by synchrotron small-angle X-ray scattering. *J. Mol. Biol.* **1996**, *262*, 559-574.

35. Tcherkasskaya, O.; Uversky, V. N. Denatured collapsed states in protein folding: example of apomyoglobin. *Proteins* **2001**, *44*, 244-254.

36. Uversky, V. N.; Ptitsyn, O. B. Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.* **1996**, *255*, 215-228.

37. Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573-6582.

38. Dunker, A. K.; Brown, C. J.; Obradovic, Z. Identification and functions of usefully disordered proteins. *Adv. Protein Chem.* **2002**, *62*, 25-49.

39. Ibers, J. A.; Holm, R. H. Modeling coordination sites in metallobiomolecules. *Science* **1980**, *209*, 223-235.

40. Tainer, J. A.; Roberts, V. A.; Getzoff, E. D. Protein metal-binding sites. *Curr. Opin. Biotechnol.* **1992**, *3*, 378-387.

41. Haase, H.; Maret, W. Fluctuations of cellular, available zinc modulate insulin signaling via inhibition of protein tyrosine phosphatases. *J. Trace Elem. Med. Biol.* **2005**, *19*, 37-42.

42. Franco, F. J.; Diaz, C.; Barcia, M.; Freire, M. Thymosin α1 is a native peptide in several tissues. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1992**, *1120*, 43-48.

43. Uversky, V. N.; Gillespie, J. R.; Millett, I. S.; Khodyakova, A. V.; Vasilenko, R. N.; Vasiliev, A. M.; Rodionov, I. L.; Kozlovskaya, G. D.; Dolgikh, D. A.; Fink, A. L.; Doniach, S.; Permyakov, E. A.; Abramov, V. M. Zn2+-Mediated Structure Formation and Compaction of the "Natively Unfolded" Human Prothymosin α. *Biochem. Biophys. Res. Commun.* **2000**, *267*, 663-668.

44. Bisaglia, M.; Tessari, I.; Mammi, S.; Bubacco, L. Interaction between alpha-synuclein and metal ions, still looking for a role in the pathogenesis of Parkinson's disease. *Neuromolecular Med.* **2009**, *11*, 239-251.

45. Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* **2005**, *579*, 3346-3354.

46. Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* **2011**, *21*, 432-440.

47. Hammoudeh, D. I.; Follis, A. V.; Prochownik, E. V.; Metallo, S. J. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J. Am. Chem. Soc.* **2009**, *131*, 7390-7401.

48. Lamberto, G. R.; Binolfi, A.; Orcellet, M. L.; Bertoncini, C. W.; Zweckstetter, M.; Griesinger, C.; Fernandez, C. O. Structural and mechanistic basis behind the inhibitory interaction of PcTS on alpha-synuclein amyloid fibril formation. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21057-21062.

49. Zahn, R.; Liu, A.; Luhrs, T.; Riek, R.; von Schroetter, C.; Lopez Garcia, F.; Billeter, M.; Calzolai, L.; Wider, G.; Wuthrich, K. NMR solution structure of the human prion protein. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 145-150.

50. Brand, I. A.; Heinickel, A.; Kratzin, H.; Soling, H. D. Properties of a 19-kDa Zn2+-binding protein and sequence of the Zn2+-binding domains. *Eur. J. Biochem.* **1988**, *177*, 561-568.

51. Gast, K.; Damaschun, H.; Eckert, K.; Schulze-Forster, K.; Maurer, H. R.; Mueller-Frohne, M.; Zirwer, D.; Czarnecki, J.; Damaschun, G. Prothymosin .alpha.: A Biologically Active Protein with Random Coil Conformation. *Biochemistry* **1995**, *34*, 13211-13218.

52. Makarova, T.; Grebenshikov, N.; Egorov, C.; Vartapetian, A.; Bogdanov, A. Prothymosin α is an evolutionary conserved protein covalently linked to a small RNA. *FEBS Lett.* **1989**, *257*, 247-250.

53. Gómez-Márquez, J.; Segade, F. Prothymosin α is a nuclear protein. *FEBS Lett.* **1988**, *226*, 217-219.

54. Haritos, A. A.; Goodall, G. J.; Horecker, B. L. Prothymosin alpha: isolation and properties of the major immunoreactive form of thymosin alpha 1 in rat thymus. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 1008-1011.

55. Hannappel, E.; Huff, T. The thymosins. Prothymosin alpha, parathymosin, and beta-thymosins: structure and function. *Vitam. Horm.* **2003**, *66*, 257-296.

56. Chen, C.; Li, M.; Yang, H.; Chai, H.; Fisher, W.; Yao, Q. Roles of thymosins in cancers and other organ systems. *World J. Surg.* **2005**, *29*, 264-270.

57. Pineiro, A.; Cordero, O. J.; Nogueira, M. Fifteen years of prothymosin alpha: contradictory past and new horizons. *Peptides* **2000**, *21*, 1433-1446.

58. Rubtsov, Y. P.; Zolotukhin, A. S.; Vorobjev, I. A.; Chichkova, N. V.; Pavlov, N. A.; Karger, E. M.; Evstafieva, A. G.; Felber, B. K.; Vartapetian, A. B. Mutational analysis of human prothymosin alpha reveals a bipartite nuclear localization signal. *FEBS Lett.* **1997**, *413*, 135-141.

59. Uversky, V. N.; Gillespie, J. R.; Millett, I. S.; Khodyakova, A. V.; Vasiliev, A. M.; Chernovskaya, T. V.; Vasilenko, R. N.; Kozlovskaya, G. D.; Dolgikh, D. A.; Fink, A. L.; Doniach, S.; Abramov, V. M. Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. *Biochemistry* **1999**, *38*, 15009-15016.

60. Yi, S.; Boys, B. L.; Brickenden, A.; Konermann, L.; Choy, W. Y. Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch. *Biochemistry* **2007**, *46*, 13120-13130.

61. Smith, M. R. Prothymosin alpha: in search of a function. *Leuk. Lymphoma* **1995**, *18*, 209-214.

62. Bachmair, A.; Finley, D.; Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **1986**, *234*, 179-186.

63. Karetsou, Z.; Sandaltzopoulos, R.; Frangou-Lazaridis, M.; Lai, C. Y.; Tsolas, O.; Becker, P. B.; Papamarcaki, T. Prothymosin alpha modulates the interaction of histone H1 with chromatin. *Nucleic Acids Res.* **1998**, *26*, 3111-3118.

64. Mosoian, A. Intracellular and extracellular cytokine-like functions of prothymosin alpha: implications for the development of immunotherapies. *Future Med. Chem.* **2011**, *3*, 1199-1208.

65. Rodriguez, P.; Vinuela, J. E.; Alvarez-Fernandez, L.; Buceta, M.; Vidal, A.; Dominguez, F.; Gomez-Marquez, J. Overexpression of prothymosin alpha accelerates proliferation and retards differentiation in HL-60 cells. *Biochem. J.* **1998**, *331 (Pt 3)*, 753-761.

66. Tsitsiloni, O. E.; Stiakakis, J.; Koutselinis, A.; Gogas, J.; Markopoulos, C.; Yialouris, P.; Bekris, S.; Panoussopoulos, D.; Kiortsis, V.; Voelter, W. Expression of alpha-thymosins in human tissues in normal and abnormal growth. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 9504-9507.

67. Tsitsilonis, O. E.; Bekris, E.; Voutsas, I. F.; Baxevanis, C. N.; Markopoulos, C.; Papadopoulou, S. A.; Kontzoglou, K.; Stoeva, S.; Gogas, J.; Voelter, W.; Papamichail, M. The prognostic value of alpha-thymosins in breast cancer. *Anticancer Res.* **1998**, *18*, 1501-1508.

68. Dominguez, F.; Magdalena, C.; Cancio, E.; Roson, E.; Paredes, J.; Loidi, L.; Zalvide, J.; Fraga, M.; Forteza, J.; Regueiro, B. J.; Puente, J. L. Tissue concentrations of prothymosin alpha: A novel proliferation index of primary breast cancer. *Eur. J. Cancer* **1993**, *29*, 893-897.

69. Haritos, A. A.; Goodall, G. J.; Horecker, B. L. Prothymosin alpha: isolation and properties of the major immunoreactive form of thymosin alpha 1 in rat thymus. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 1008-1011.

70. Karetsou, Z.; Kretsovali, A.; Murphy, C.; Tsolas, O.; Papamarcaki, T. Prothymosin alpha interacts with the CREB-binding protein and potentiates transcription. *EMBO Rep.* **2002**, *3*, 361-366.

71. Mosoian, A.; Teixeira, A.; Burns, C. S.; Khitrov, G.; Zhang, W.; Gusella, L.; Klotman, P.; Klotman, M. Influence of prothymosin-alpha on HIV-1 target cells. *Ann. N. Y. Acad. Sci.* **2007**, *1112*, 269-285.

72. Mosoian, A.; Teixeira, A.; Burns, C. S.; Sander, L. E.; Gusella, G. L.; He, C.; Blander, J. M.; Klotman, P.; Klotman, M. E. Prothymosin-alpha inhibits HIV-1 via Toll-like receptor 4-mediated type I interferon induction. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10178-10183.

73. Mosoian, A.; Teixeira, A.; High, A. A.; Christian, R. E.; Hunt, D. F.; Shabanowitz, J.; Liu, X.; Klotman, M. Novel function of prothymosin alpha as a potent inhibitor of human

immunodeficiency virus type 1 gene expression in primary macrophages. *J. Virol.* **2006**, *80*, 9200-9206.

74. Dong, G.; Callegari, E. A.; Gloeckner, C. J.; Ueffing, M.; Wang, H. Prothymosin-alpha interacts with mutant huntingtin and suppresses its cytotoxicity in cell culture. *J. Biol. Chem.* **2012**, *287*, 1279-1289.

75. Chichkova, N. V.; Evstafieva, A. G.; Lyakhov, I. G.; Tsvetkov, A. S.; Smirnova, T. A.; Karapetian, R. N.; Karger, E. M.; Vartapetian, A. B. Divalent metal cation binding properties of human prothymosin alpha. *Eur. J. Biochem.* **2000**, *267*, 4745-4752.

76. Suhasini, M.; Reddy, T. R. Cellular proteins and HIV-1 Rev function. *Curr. HIV. Res.* **2009**, *7*, 91-100.

77. Sundaram, M.; Saghayam, S.; Priya, B.; Venkatesh, K. K.; Balakrishnan, P.; Shankar, E. M.; Murugavel, K. G.; Solomon, S.; Kumarasamy, N. Changes in antioxidant profile among HIV-infected individuals on generic highly active antiretroviral therapy in southern India. *Int. J. Infect. Dis.* **2008**, *12*, e61-6.

78. Corbeil, J.; Sheeter, D.; Genini, D.; Rought, S.; Leoni, L.; Du, P.; Ferguson, M.; Masys, D. R.; Welsh, J. B.; Fink, J. L.; Sasik, R.; Huang, D.; Drenkow, J.; Richman, D. D.; Gingeras, T. Temporal gene regulation during HIV-1 infection of human CD4+ T cells. *Genome Res.* **2001**, *11*, 1198-1204.

79. Cota, M.; Mengozzi, M.; Vicenzi, E.; Panina-Bordignon, P.; Sinigaglia, F.; Transidico, P.; Sozzani, S.; Mantovani, A.; Poli, G. Selective inhibition of HIV replication in primary macrophages but not T lymphocytes by macrophage-derived chemokine. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9162-9167.

80. Uversky, V. N.; Karnoup, A. S.; Khurana, R.; Segel, D. J.; Doniach, S.; Fink, A. L. Association of partially-folded intermediates of staphylococcal nuclease induces structure and stability. *Protein Sci.* **1999**, *8*, 161-173.

81. Sattler, M.; Schleucher, J.; Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93-158.

82. Alexandrescu, A. T.; Shortlet, D. Backbone Dynamics of a Highly Disordered 131 Residue Fragment of Staphylococcal Nuclease. *J. Mol. Biol.* **1994**, *242*, 527-546.

83. Stanfield, R. L.; Wilson, I. A. Protein-peptide interactions. *Curr. Opin. Struct. Biol.* **1995**, *5*, 103-113.

84. Kubota, S.; Adachi, Y.; Copeland, T. D.; Oroszlan, S. Binding of human prothymosin alpha to the leucine-motif/activation domains of HTLV-I Rex and HIV-1 Rev. *Eur. J. Biochem.* **1995**, *233*, 48-54.

85. Kjems, J.; Askjaer, P. In *Rev protein and its cellular partners;* Advances in Pharmacology; Academic Press: Vol. Volume 48, pp 251-298.

86. Wilson, C. L.; Monteith, W. B.; Danell, A. S.; Burns, C. S. Purification and characterization of the central segment of prothymosin-alpha: methodology for handling highly acidic peptides. *J. Pept. Sci.* **2006**, *12*, 721-725.

87. El Alaoui, E. A.; Champmartin, D.; Rubini, P. Complexes of EDTA in aqueous solutions. Structural aspects from a ^{13}C NMR relaxation study. *J. Chem. Soc. , Dalton Trans.* **2001**, 2153-2156.

88. Kent, S. B. Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* **1988**, *57*, 957-989.

89. Kates, ,StevenA.; Albericio, ,Fernando Rearrangement of Glu(OtBu)- and Asp(OtBu)-containing peptides upon fluoride treatment in solid-phase synthesis. *Let. in pep. sci.* **1994**, 1, 213-220.

90. Tam, J. P.; Riemen, M. W.; Merrifield, R. B. Mechanisms of aspartimide formation: the effects of protecting groups, acid, base, temperature and time. *Pept. Res.* **1988**, *1*, 6-18.

91. Palasek, S. A.; Cox, Z. J.; Collins, J. M. Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis. *J. Pept. Sci.* **2007**, *13*, 143-148.

92. Prestidge, R. L.; Harding, D. R.; Hancock, W. S. Use of substituted benzyl esters as carboxyl-protecting groups in solid-phase peptide synthesis. *J. Org. Chem.* **1976**, *41*, 2579-2583.

93. Rodthongkum, N.; Chen, Y.; Thayumanavan, S.; Vachet, R. W. Selective enrichment and analysis of acidic peptides and proteins using polymeric reverse micelles and MALDI-MS. *Anal. Chem.* **2010**, *82*, 8686-8691.

94. - Bratby, D. M.; - Coyle, S.; - Gregson, R. P.; - Hardy, G. W.; - Young, G. T. - Amino-acids and peptides. Part 42. Synthesis of a protected docosa peptide having the sequence of mast-cell degranulating peptide. *J. Chem. Soc., Perkin Trans.* **1979,** 1 , 1901-1907.

95. Garner, R.; Young, G. T. Amino-acids and peptides. Part XXXIII. Synthesis of Val5-angiotensin-II by the picolyl ester method. *J. Chem. Soc. C*, **1971,** *1*, 50-53.

96. CAMBLE, ,R.; GARNER, ,R.; YOUNG, G.,T. Novel Facilitation of Peptide Synthesis. *Lett. to nature*. **1968**, 217, 247-248.

97. Rizo, J. Albericio. F , Romero. G, Garcia-Echeverria. C , Claret. J, Carlos Muller , Ernest Giralt , Enrique Pedroso. Use of polar picolyl protecting groups in peptide synthesis. *J. Org. Chem.*, **1988**, 53 5386–5389.

99. Kelly, S. M.; Price, N. C. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **2000**, *1*, 349-384.

100. Clarke, D. T.; Doig, A. J.; Stapley, B. J.; Jones, G. R. The α-helix folds on the millisecond time scale. *Proceedings of the National Academy of Sciences* **1999**, *96*, 7232-7237.

101. CHAFFOTTE, A. F.; GUILLOU, Y.; GOLDBERG, M. E. Inclusion bodies of the thermophilic endoglucanase D from Clostridium thermocellum are made of native enzyme that resists 8 M urea. *European Journal of Biochemistry* **1992**, *205*, 369-373.

102. Greenfield, N. J. Applications of circular dichroism in protein and peptide analysis. *TrAC Trends in Analytical Chemistry* **1999**, *18*, 236-244.

103. Strickland, E. H. Aromatic contributions to circular dichroism spectra of proteins. *CRC Crit. Rev. Biochem.* **1974**, *2*, 113-175.

104. Uversky, V. N. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J.* **2009**, *28*, 305-325.

105. Uversky, V. N.; Li, J.; Fink, A. L. Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. A possible molecular NK between Parkinson's disease and heavy metal exposure. *J. Biol. Chem.* **2001**, *276*, 44284-44296.

106. Uversky, V. N.; Li, J.; Fink, A. L. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* **2001**, *276*, 10737-10744.

107. Li, J.; Uversky, V. N.; Fink, A. L. Conformational behavior of human alpha-synuclein is modulated by familial Parkinson's disease point mutations A30P and A53T. *Neurotoxicology* **2002**, *23*, 553-567.

108. Uversky, V. N.; Li, J.; Bower, K.; Fink, A. L. Synergistic effects of pesticides and metals on the fibrillation of alpha-synuclein: implications for Parkinson's disease. *Neurotoxicology* **2002**, *23*, 527-536.

109. Uversky, V. N.; Permyakov, S. E.; Zagranichny, V. E.; Rodionov, I. L.; Fink, A. L.; Cherskaya, A. M.; Wasserman, L. A.; Permyakov, E. A. Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: a natively unfolded protein. *J. Proteome Res.* **2002**, *1*, 149-159.

110. Permyakov, S. E.; Millett, I. S.; Doniach, S.; Permyakov, E. A.; Uversky, V. N. Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. *Proteins: Structure, Function, and Bioinformatics* **2003**, *53*, 855-n/a.

111. Konno, T.; Tanaka, N.; Kataoka, M.; Takano, E.; Maki, M. A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. *Biochim. Biophys. Acta* **1997**, *1342*, 73-82.

112. Lynn, A.; Chandra, S.; Malhotra, P.; Chauhan, V. S. Heme binding and polymerization by Plasmodium falciparum histidine rich protein II: influence of pH on activity and conformation. *FEBS Lett.* **1999**, *459*, 267-271.

113. Johansson, J.; Gudmundsson, G. H.; Rottenberg, M. E.; Berndt, K. D.; Agerberth, B. Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37. *J. Biol. Chem.* **1998**, *273*, 3718-3724.

114. Dudev, T.; Lim, C. Principles governing Mg, Ca, and Zn binding and selectivity in proteins. *Chem. Rev.* **2003**, *103*, 773-788.

115. Dudev, T.; Lim, C. Metal Binding and Selectivity in Zinc Proteins. *J. Chin. Chem. Soc.* **2003**, *50*, 1093-1102.

116. Kjaergaard, M.; Norholm, A. B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B. Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? *Protein Sci.* **2010**, *19*, 1555-1564.

117. Uversky, V. N.; Narizhneva, N. V. Effect of natural ligands on the structural properties and conformational stability of proteins. *Biochemistry (Mosc)* **1998**, *63*, 420-433.

118. Uversky, V. N. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J.* **2009**, *28*, 305-325.

119. Privalov, P. L.; Potekhin, S. A. Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol.* **1986**, *131*, 4-51.

120. Even, J.; Bertault, M.; Girard, A.; Délugeard, Y.; Marqueton, Y. Optical and calorimetric studies on the role of lattice mode softening in assisting a thermally enhanced solid state reaction. *Chemical Physics Letters* **1997**, *267*, 585-589.

121. Lin, L. N.; Mason, A. B.; Woodworth, R. C.; Brandts, J. F. Calorimetric studies of the N-terminal half-molecule of transferrin and mutant forms modified near the Fe(3+)-binding site. *Biochem. J.* **1993**, *293 (Pt 2)*, 517-522.

122. Ladbury, J. E.; Chowdhry, B. Z. Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem. Biol.* **1996**, *3*, 791-801.

123. Stockar, U.; Marison, I. In *The use of calorimetry in biotechnology;* Springer Berlin Heidelberg: 1989; Vol. 40, pp 93-136.

124. Weber, P. C.; Salemme, F. R. Applications of calorimetric methods to drug discovery and the study of protein interactions. *Curr. Opin. Struct. Biol.* **2003**, *13*, 115-121.

125. Varghese, N.; Vivekchand, S. R. C.; Govindaraj, A.; Rao, C. N. R. A calorimetric investigation of the assembly of gold nanorods to form necklaces. *Chemical Physics Letters* **2008**, *450*, 340-344.

126. Schick, C.; Lexa, D.; Leibowitz, L. *Differential Scanning Calorimetry and Differential Thermal Analysis;* Characterization of Materials; John Wiley & Sons, Inc. 2002.

127. Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nöllmann, M. Heat does not come in different colours: entropy–enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.* **2001**, *93*, 215-230.

128. Privalov, P. L. In *Protein?DNA Interaction;* Microcalorimetry of Macromolecules; John Wiley & Sons, Inc.: 2012; pp 297-351.

129. Cooper, A. Heat capacity of hydrogen-bonded networks: an alternative view of protein folding thermodynamics. *Biophys. Chem.* **2000**, *85*, 25-39.

130. Gill, P.; Moghadam, T. T.; Ranjbar, B. Differential scanning calorimetry techniques: applications in biology and nanoscience. *J. Biomol. Tech.* **2010**, *21*, 167-193.

131. McElhaney, R. N. The use of differential scanning calorimetry and differential thermal analysis in studies of model and biological membranes. *Chem. Phys. Lipids* **1982**, *30*, 229-259.

132. Biltonen, R. L.; Lichtenberg, D. The use of differential scanning calorimetry as a tool to characterize liposome preparations. *Chem. Phys. Lipids* **1993**, *64*, 129-142.

133. Giron, D. Applications of Thermal Analysis and Coupled Techniques in Pharmaceutical Industry. *Journal of Thermal Analysis and Calorimetry* **2002**, *68*, 335-357.

134. Demetzos, C. Differential Scanning Calorimetry (DSC): a tool to study the thermal behavior of lipid bilayers and liposomal stability. *J. Liposome Res.* **2008**, *18*, 159-173.

135. Coleman, N. J.; Craig, D. Q. M. Modulated temperature differential scanning calorimetry: A novel approach to pharmaceutical thermal analysis. *Int. J. Pharm.* **1996**, *135*, 13-29.

136. Privalov, P. L.; Dragan, A. I. Microcalorimetry of biological macromolecules. *Biophys. Chem.* **2007**, *126*, 16-24.

137. Chen, T.; Oakley, D. M. Thermal analysis of proteins of pharmaceutical interest. *Thermochimica Acta* **1995**, *248*, 229-244.

138. Sturtevant, J. M. Biochemical Applications of Differential Scanning Calorimetry. *Annu. Rev. Phys. Chem.* **1987**, *38*, 463-488.

139. Cooper, M. A. Label-free screening of bio-molecular interactions. *Anal. Bioanal Chem.* **2003**, *377*, 834-842.

140. Fisher, H. F.; Singh, N. Calorimetric methods for interpreting protein-ligand interactions. *Methods Enzymol.* **1995**, *259*, 194-221.

141. Minetti, C. A.; Remeta, D. P. Energetics of membrane protein folding and stability. *Arch. Biochem. Biophys.* **2006**, *453*, 32-53.

142. Shnyrov, V. L.; Sanchez-Ruiz, J. M.; Boiko, B. N.; Zhadan, G. G.; Permyakov, E. A. Applications of scanning microcalorimetry in biophysics and biochemistry. *Thermochimica Acta* **1997**, *302*, 165-180.

143. Lucio, M.; Lima, J. L.; Reis, S. Drug-membrane interactions: significance for medicinal chemistry. *Curr. Med. Chem.* **2010**, *17*, 1795-1809.

144. Mouritsen, O. G.; Bloom, M. Models of lipid-protein interactions in membranes. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 145-171.

145. Cueto, M.; Dorta, M. J.; Munguia, O.; Llabres, M. New approach to stability assessment of protein solution formulations by differential scanning calorimetry. *Int. J. Pharm.* **2003**, *252*, 159-166.

146. Conejero-Lara, F.; Sanchez-Ruiz, J. M.; Mateo, P. L.; Burgos, F. J.; Vendrell, J.; Aviles, F. X. Differential scanning calorimetric study of carboxypeptidase B, procarboxypeptidase B and its globular activation domain. *Eur. J. Biochem.* **1991**, *200*, 663-670.

147. Yamagata, Y.; Kubota, M.; Sumikawa, Y.; Funahashi, J.; Takano, K.; Fujii, S.; Yutani, K. Contribution of hydrogen bonds to the conformational stability of human lysozyme: calorimetry and X-ray analysis of six tyrosine --> phenylalanine mutants. *Biochemistry* **1998**, *37*, 9355-9362.

148. Haynie, D. T.; Freire, E. Estimation of the folding/unfolding energetics of marginally stable proteins using differential scanning calorimetry. *Anal. Biochem.* **1994**, *216*, 33-41.

149. Plotnikov, V. V.; Brandts, J. M.; Lin, L. N.; Brandts, J. F. A new ultrasensitive scanning calorimeter. *Anal. Biochem.* **1997**, *250*, 237-244.

150. Cooper, A. Protein Heat Capacity: An Anomaly that Maybe Never Was. *J. Phys. Chem. Lett.* **2010**, *1*, 3298-3304.

151. Scholtz, J. M.; Baldwin, R. L. The mechanism of alpha-helix formation by peptides. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95-118.

152. Kosol, S.; Contreras-Martos, S.; Cedeno, C.; Tompa, P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **2013**, *18*, 10802-10828.

153. Latypov, R. F.; Harvey, T. S.; Liu, D.; Bondarenko, P. V.; Kohno, T.; Fachini II, R. A.; Rosenfeld, R. D.; Ketchem, R. R.; Brems, D. N.; Raibekas, A. A. Biophysical Characterization of Structural Properties and Folding of Interleukin-1 Receptor Antagonist. *J. Mol. Biol.* **2007**, *368*, 1187-1201.

154. Kowalewski, J.; Mäler, L. *Nuclear spin relaxation in liquids: theory, experiments, and applications;* Series in chemical physics; Taylor & Francis: New York, 2006; Vol. 2, pp 426.

155. Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **1990**, *89*, 496-514.

156. Zarrinpar, A.; Bhattacharyya, R. P.; Lim, W. A. The Structure and Function of Proline Recognition Domains. *Sci. STKE* **2003**, *2003*, re8.

157. Schnell, J. R.; Dyson, H. J.; Wright, P. E. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 119-140.

158. Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y. X.; Kustanovich, I.; Louis, J. M.; Torchia, D. A. Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci.* **2002**, *11*, 221-232.

159. Pasat, G.; Zintsmaster, J. S.; Peng, J. W. Direct 13C-detection for carbonyl relaxation studies of protein dynamics. *Journal of Magnetic Resonance* **2008**, *193*, 226-232.

160. Michael, S. F.; Kilfoil, V. J.; Schmidt, M. H.; Amann, B. T.; Berg, J. M. Metal binding and folding properties of a minimalist Cys2His2 zinc finger peptide. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 4796-4800.

161. Garapati, S.; Burns, C. S. The 4-pyridylmethyl ester as a protecting group for glutamic and aspartic acids: ?flipping? peptide charge states for characterization by positive ion mode ESI-MS. *Journal of Peptide Science* **2014**, *20*, 191-195.

162.Dudev, T.; Lim, C. Metal binding affinity and selectivity in metalloproteins: Insights from computational studies. Annu. Rev. Biophys. **2008**, *37*, 97-116.

163. Srikanth, R.; Wilson, J.; Burns, C. S.; Vachet, R. W. Identification of the Copper(II) Coordinating Residues in the Prion Protein by Metal-Catalyzed Oxidation Mass

Spectrometry: Evidence for Multiple Isomers at Low Copper(II) Loadings. *Biochemistry* **2008**, *47*, 9258-9268.

164. Hong, L.; Carducci, T. M.; Bush, W. D.; Dudzik, C. G.; Millhauser, G. L.; Simon, J. D. Quantification of the binding properties of Cu2+ to the amyloid beta peptide: coordination spheres for human and rat peptides and implication on Cu2+-induced aggregation. *J. Phys. Chem. B* **2010**, *114*, 11261-11271.